Università degli Studi di Napoli "Federico II"

Dipartimento di Scienze Fisiche

Dottorado di Ricerca in Fisica Fondamentale ed Applicata

Search for Higgs-like resonances in the $ZZ \to l^{\pm}l^{\pm}q\bar{q}$ decay channel with the ATLAS experiment at LHC

Thesis for Doctor of Philosophy degree submitted by

Arturo Sánchez Pineda

Advisors:

Proff. Mariagrazia Alviggi

Dott. Francesco Conventi

Coordinator:

Prof. Raffaele Velotta

XXV ciclo 2010/2013

If I could remember the names of all these particles, I'd be a botanist...
Enrico Fermi

Dedication

To my moms, my strengths of will... A mi madre y mi mamita... mis fuerzas de voluntad

Acknowledgments

Have been a lot of people that in one way or in other helped me to arrived at this point. My family in Táchira, who despite the distance, they have always been there and sending their prayers and energies. My professors in Venezuela, Luís, Alejandra, Nelsón, Chourio, Fulgencio, Victor, Patricia, Domingo, who with great ethics and passion for knowledge guide to me, specially to Professor Gianfranco Spavieri, whom I consider my friend and mentor. My family in Mérida who always had a word of encouragement and appreciation, and very good conversations. To my friends, "mi patrulla" that was always there, frank but true. To the Rivas Family, for their invaluable help and passion. To Estefanía, who has been with my Mother are my reason to exist. Continue like this! My godmother (mi Madrina), who always protected me as a child and advise. To Orlando for their advice and expertise, to Roman for much help and words of encouragement. To Reina for being my guide throughout this journey, really the person I most admire. To Fidelina, who never left me faint. "Mamita, millones de gracias por haber sido y ser mi mejor amiga." To my Grandmothers, the stronger women that I never knew. Gracias mamá Ismelda donde quiera q estés. To Diegito, Ruben Dario, Cesar, William, you are my friends! A Sandro de Cecco and José Ocáriz for their help, energy and honest advices. To the ATLAS teams in Naples and Rome, really my admiration and respect for all. To Gianpaolo, Prof. Canale, Elvira, Givi, Alessandra, Massimo, Paolo, Vincenzo, Marco, Valerio, Francesco, Alessia, Mirko, Giovanni and other member of that team for their professionalism and humanity. To Prof. Mariagrazia Alvggi and Francesco Conventi for all time, help, patience and knowledge that I received. To Prof. Marrucci for his great welcome to Naples. A Barbara, so excellent friend and confident!. To Laura M., who changed my outlook, strong and cheerful flower!. To all the rest of the people who gave me the knowledge, strong, passion and energy to complete this unique period in my life. Gracias!. A mi mamá Xiomara, jamás han sido escritas las palabras que logren abarcar el amor, respeto, admiración y agradecimiento que siento por ella.

Abstract

The search for the Standard Model (SM) Higgs boson is one of the most crucial goals of the LHC physics program. The high center-of-mass energy ($\sqrt{s}$) of the LHC enables not just to search for the SM Higgs boson at low mass, i.e. between 120 GeV $- 2 \times m_Z$, but to extend it to much larger masses, in the range 200 GeV $-$ 1 TeV. Although a large portion of last mass range is indirectly excluded at 95% Confidence Level (CL) by global fits to SM observables, it is crucial to complement such indirect limits by direct searches. Further, possible extensions to the SM can conspire to allow a heavy Higgs boson to be compatible with existing measurements and latest Higgs boson-candidate discovery. My Ph.D. research activity was focused in the $H \to ZZ \to l^{\pm}l^{\pm}q\bar{q}$ analysis in the full Higgs mass range, which has been split in: Low Mass (120 GeV $- 2 \times m_Z$) and High Mass (200 GeV $-$ 1 TeV), where the crucial work was performed in the inclusion, for the first time into the ATLAS research program, of the study of the $H \to ZZ(^*) \to l^{\pm}l^{\pm}q\bar{q}$ decay in the Low Mass range using 2011 data recorded at $\sqrt{s} = 7$ TeV, principal subject of this dissertation. Recently, taking the mentioned discover of a particle compatible with the Higgs boson (*July 4th, 2012*) like a strong guide to continue this research using the full available 2012 recorded data at $\sqrt{s} = 8$ TeV, and knowing that several Beyond SM models (BSM), compatible with the observed ~ 125 GeV resonance ($h1$) and Electroweak (EW) fit, predict a second 'SM-like' heavy Higgs state $h2$, it will be possible to search for SM resonance in the heavy mass region (400 GeV $-$ 1 TeV) looking for excess with respect to the SM predictions regardless which model could produce such excess. This is our actual research activity into the ATLAS experiment and a view of its state of the art is included into this dissertation, as the novel tools developed during such a studies.

Introduction

The most complex machine in the World, the Large Hadron Collider (LHC) is starting its first long shutdown after three very fruitful years of scientific program.

Conceived and constructed with the finality to recreate the initial conditions of the Universe and to test the Standard Model (SM) of particle physics, the most successful theory of the modern physics, responsible to explain the behaviour of the elementary particles and their interactions.

Since the predictions of the SM have been tested in many aspects (and by many experiments), until some months ago remained $-$*apparently*$-$ without experimental evidence of one of its fundamental predictions. The mechanism responsible for the electroweak (EW) symmetry breaking, that incorporates masses to the particles and accurately predicts the observed interactions between them, where in its simplest version, predicts a scalar boson that could be maybe discovered last year at CERN.

This particle, called the Higgs boson[10] knowing for been the last undiscovered particle in the SM, and the subject of this dissertation, have been searched mainly for the two biggest experiments at LHC, ATLAS and CMS, that, using data collected during the 2011 ($4.8fb^{-1}$) and the first part of 2012 ($5.8fb^{-1}$) at center of mass energy of $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV respectively, report the observation of a new particle, a boson, with a mass close to 126 GeV. The discovery of a Higgs boson-like particle have open the door not just to the SM ultimate test, but also for the possibility to exclude $-$or not$-$ theories Beyond the Standard Model (BSM).

After July 4^{th} 2012 news [23], and due to the knowledge that the Higgs mechanism (and its corresponding boson) is not the only theory of physics that can alter the electroweak symmetry and produce different distinguish particles, as Supersymmetry [33, 34] that predicts a minimum of five scalar an pseudoscalar particles in place of the Higgs boson, or the two-Higgs-doublet (2HDM)[31] models require that an additional Higgs doublet be added like is done in the Minimal Supersymmetric Standard Model (MSSM)[32].

Added to the Higgs hunting, my PhD thesis reports on the search for de Higgs boson in $4.7fb^{-1}$ proton-proton collision data recoded during 2011. The data were collected by the ATLAS detector. The focus is on the $H \to ZZ \to l^{\pm}l^{\pm}q\bar{q}$ decay mode and the analysis uses cut-based techniques.

Backgrounds are estimated using methods relying on both data and Monte Carlo simulations, apart of the development of new techniques to improve the rejection of such a background, before the Confidence Level (CL) method of statistical fitting is used to assess the significance of the observed results.

This thesis uses the so-called natural units, a convention commonly employed by the experimental high energy physics community where $c = h/2\pi = 1$, where c is the speed of light and h is the Planck's constant. Therefore, masses and momenta are given in GeV rather than GeV/c^2 and GeV/c, respectively.

This thesis is structured as follows: after a brief introduction on the theoretical framework of the SM of particle physics adding theoretical and experimental constraints for the mass of the Higgs boson −including the latest LHC results−, plus a view into Beyond SM scenarios is done in Chapter 1, the LHC complex at CERN and ATLAS detector are described in Chapters 2.

The Chapter 3 describes the reconstruction, identification and calibration of physics objects combining the information of the ATLAS sub-detectors. Only those physics objects particularly relevant for this thesis are described: electrons, muons, jets and the missing transverse energy.

The Chapter 4 describes the main Higgs searches in ATLAS split into the so-called Low Mass range and High Mass range. The latest combination results have been included.

In the Chapter 5 a complete overview of the $H \to ZZ(^*) \to l^{\pm}l^{\pm}q\bar{q}$ decay modes is done. Into this chapter the description of the MC −signal and background− processes and the data used along this thesis is done, following for the object, tools and event selection of the $H \to ZZ(^*) \to l^{\pm}l^{\pm}q\bar{q}$ low mass analysis, where was included novel procedures −for such a range of masses− on the multi-jet background estimation and the jet paring selection. A *state of the art* of its equivalent High Mass analysis is included.

The Chapters 6 correspond to the presentation of the Results of this dissertation for the $H \to ZZ \to l^{\pm}l^{\pm}q\bar{q}$ first ATLAS analysis at Low Mass range using 2011 data. A view into the status of the the 2012 Low and High Mass analyses are presented.

Inside the new tools applied along the thesis work, the optimization of signal significance using a quark/gluon separation technique is described in Chapter 7.

Finally, in Chapter 8 the Conclusions are presented. An important note, the $H \to ZZ \to l^{\pm}l^{\pm}q\bar{q}$ (high and low mass) analyses using $20.7fb^{-1}$ data collected in 2012 at center of mass energy $\sqrt{s} = 8$ TeV is still underway; therefore no final results can be presented in this document.

Contents

Chapter 1

The Standard Model and Beyond

The Standard Model (SM) of particle physics is a non-abelian gauge theory that describes the strong and electroweak interactions.

Proposed at the end of the 60's, it obtained many experimental confirmations and it is the model used currently to study the high energy physics.

This chapter summarizes the relevant theoretical aspects used into the presented analyses. Starting with a overall view of the fundamental elements of the model *-particles and forces-* arriving until a description of the Standard Model Higgs boson and the efforts to find it.

1.1 Elementary Particles

Nowadays the known physical world is described in a big proportion in terms of fundamental matter particles and their interactions thanks to the SM. It describes all known particles and three of the four known fundamental interactions (the electromagnetic, weak and strong interactions). Within the SM the particles are classified by their spin as either

- half-integer spin particles called fermions obeying Fermi-Dirac statistics. These form the matter particles.
- integer spin particles called bosons. These particles obey Bose-Einstein statistics and their exchange between the fermions describes the fundamental interactions.

1.1.1 The Fermions (Matter Particles)

The fermions are categorised into two types, the quarks and the leptons. The quarks are given a baryon number $B = \frac{1}{3}$. The leptons are assigned a lepton number $L = 1$ and do not interact strongly. Each type of fermion consists of three families or

generations, each of which consists of two distinct particles (often referred to as a doublet of particles). Figure 1.2 lists the quarks and leptons and other basic properties [5].

The first generation of quarks consists of the up (u) quark with +2 electric charge 3 (in units of electron charge, e) and the down (d) quark with -1 electric charge. The 3 other generations consist of a u-type and a d-type quark but are successively heavier than the first generation. The second generation consists of the strange (s) and charm (c) quarks while the third generation consists of the bottom (b) and top (t) quarks. Quarks carry colour charge and as such each comes in three distinct colour states (red, green or blue). Each doublet of leptons is composed of an electrically charged lepton and its corresponding neutral neutrino. As with quarks, the mass of the charged leptons in the doublet increases with generation. The first generation consists of the electron (e) and its neutrino (ν_e), the second the muon (μ) and its neutrino (ν_μ) and the third the tau (τ) and its neutrino (ν_τ). Each quark and lepton have a corresponding anti-particle, denoted with a bar. Anti-particles have opposite electric charge to the corresponding particle but the same mass. In nature quarks are only found within composite hadrons, composed of either three quarks making a baryon or in quark anti-quark states called mesons.

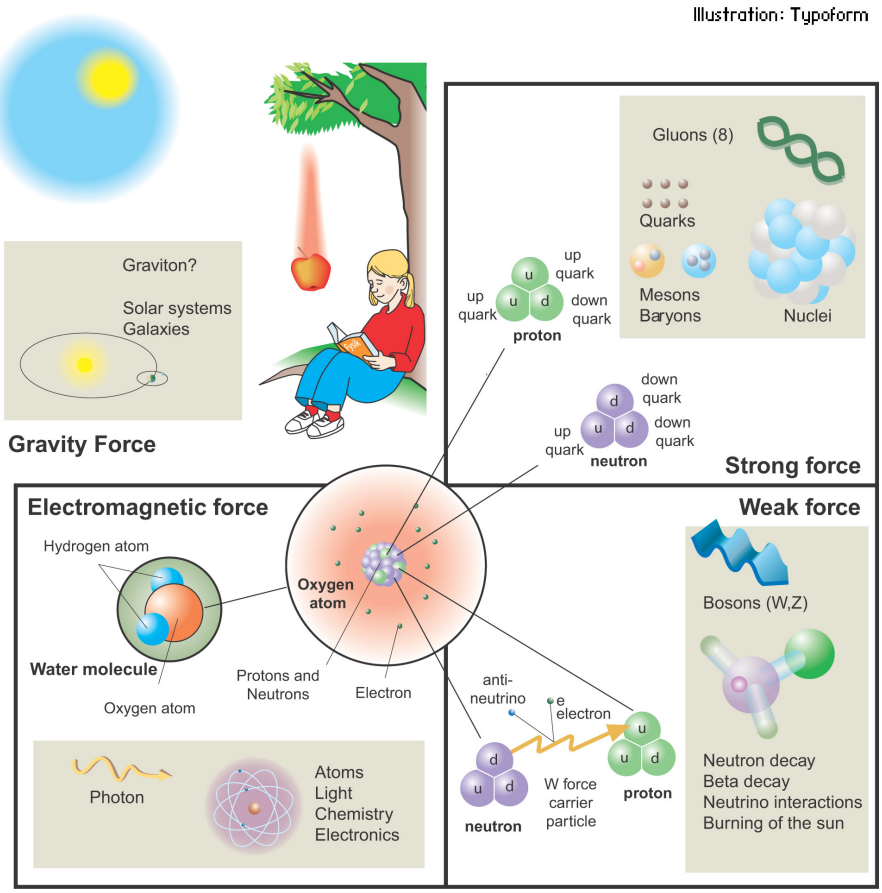


Figure 1.1 Artistic representation of the fundamental forces in Nature.

1.1.2 The Bosons (Forces and Carrier Particles)

Interactions between the fermions are mediated by the absorption and emission of integer spin particles called bosons. This gives rise to four fundamental forces, represented in the Figure 1.1. The electromagnetic force makes the electron bind to nuclei and more generally, molecule formation underpinning Chemistry. It is mediated by the photon (γ). The strong force is responsible for holding nuclei together and is mediated by eight massless gluons (g). The weak force explains decay and is mediated by exchange of W and Z bosons. Gravity is responsible for galactic formation. It is the weakest of all the forces and is negligible at the energy scales considered in particle physics.

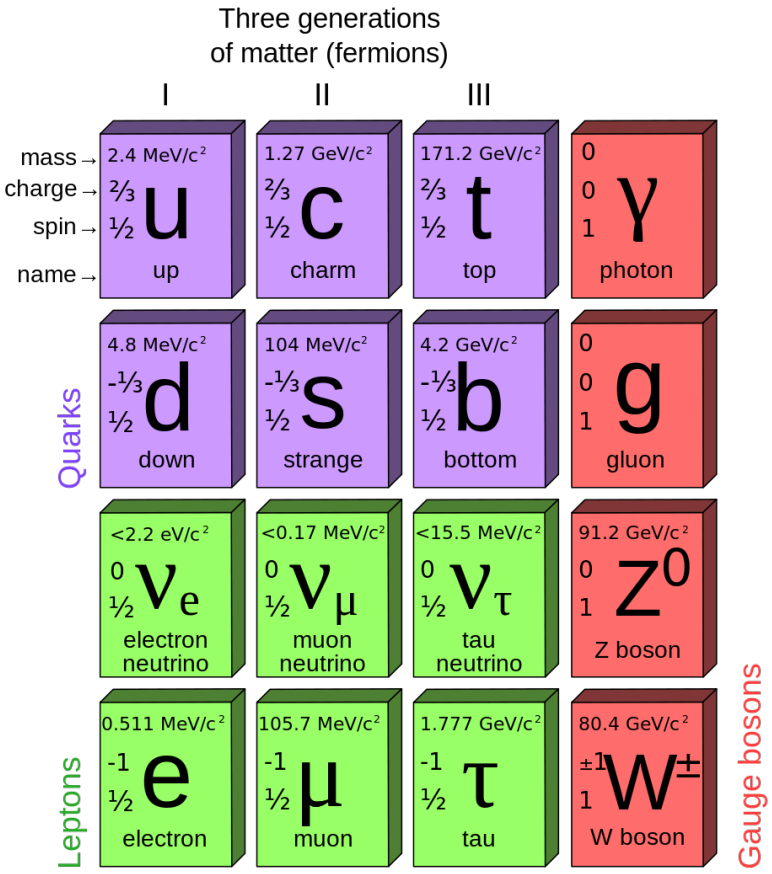


Figure 1.2 Standard Model of Elementary Particles.

1.2 The Standard Model

The SM is a theoretical framework of quantum field theory [6] in which the elementary particles are the quanta of the underlying fields and the interactions are a consequence of the principle of local gauge invariance. As yet attempts to incorporate gravity using this approach have failed.

The time-line of the SM becoming an unified theory of the forces that it describes started with the development of the quantum field theory of electromagnetic interactions, called Quantum Electrodynamics.

Subsequently in the 1960's an electroweak theory was developed unifying the electromagnetic and weak interactions.

Finally the electroweak theory was unified with the theory of the strong interactions (Quantum Chromodynamics) giving what is understood as the SM today.

1.2.1 Quantum Electrodynamics (QED)

All quantum electromagnetic interactions consist of the interaction of charged fermions with the quantum of the electromagnetic field, the photon. A basic form of such an interaction, is shown in Figure 1.3.

It shows the interaction of a charged fermions, a electron pair with a photon γ to produce a muon pair. As with all interactions, the strength is characterized by a coupling constant associated to each vertex. The electromagnetic force couples to electric charge and so this defines the strength of electromagnetic interactions. This vertex corresponds to the basic building block from which all QED processes can be represented. Complete QED processes represented in this way are called Feynman diagrams. Feynman diagrams with the smallest number of vertices for a given process to occur are referred to as tree-level or leading-order whereas diagrams with a higher number of vertices are called higher order diagrams. A detailed picture of any QED process can be obtained by summing over all possible internal states and this corresponds to summing over all Feynman diagrams of all orders.

It is convenient to use the Lagrangian formalism to describe the interactions of fermions within the SM, whereby such interactions can be described in terms of an action involving a Lagrangian acting on the fermion fields. The dynamics of non-interacting massive fermion fields is described by the Dirac equation [7] and as such the relevant Lagrangian is the Dirac Lagrangian. In 1954 Yang and Mills[8] proposed a framework for theories involving the exchange of vector bosons, such as those in QED. The Lagrangian in a Yang-Mills theory is invariant under transformations that are a function of space and time. This local gauge symmetry provides an accurate description of physical interactions and is therefore a desirable property of the SM theories. In order to ensure this symmetry is retained, a new massless vector field must be introduced which is identified as the photon field. This gives rise to the QED Lagrangian which is then observed to contain interaction terms, represented by processes such as in Figure 1.3. The requirement that the physical system remains invariant under local gauge transformations results in the conservation of electron

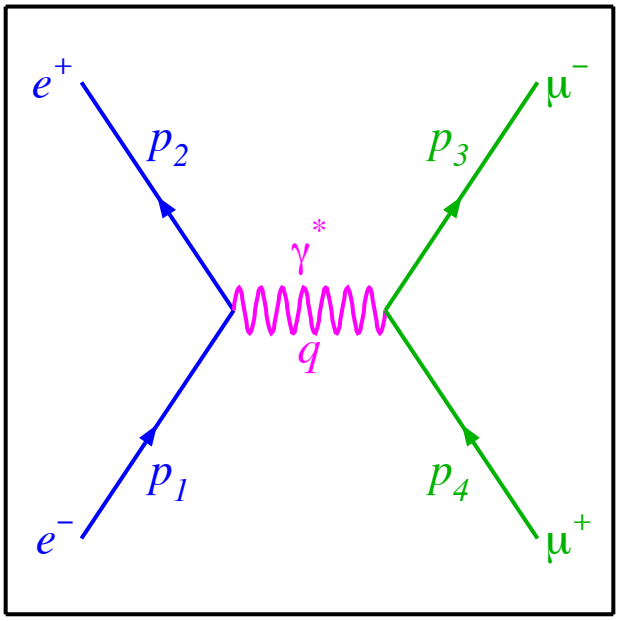


Figure 1.3 Feynman diagram of a muon pair production with a γ virtual mediator.

charge, as confirmed by experiment. A local gauge transformation in QED can be represented as a transformation under the symmetry group $U(1)$ corresponding to all unitary matrices of dimension 1×1. In this way the local gauge group of QED is called $U(1)$. Using different symmetry groups, the same principle may be extended to the strong and weak interactions.

1.2.2 Quantum Chromodynamics (QCD)

Quantum Chromodynamics (QCD) is the quantum field theory describing the strong interactions. The strong force couples to colour charge, so only the coloured gluons and quarks are involved in strong interactions. The most basic QCD interaction vertex, involving the interaction of quarks (q) with a gluon (g) is shown in Figure 1.4. The local gauge group of QCD is $SU(3)_C$ corresponding to the unitary group of 3×3 matrices with determinant 1. The three dimensional nature of this group is a consequence of there being three quark colours (C) and as such the quark fields transforming in the vector space of colour. To preserve local gauge invariance of the Dirac Lagrangian eight massless fields must be introduced which correspond to the eight gluons. These fields are also vector fields because the gluons have an intrinsic colour charge.

The $SU(3)_C$ group is an example of a non-abelian group, because its generators do not commute. This is a consequence of gluons carrying a colour charge and leads to self interaction terms in the QCD Lagrangian. This behaviour gives rise to

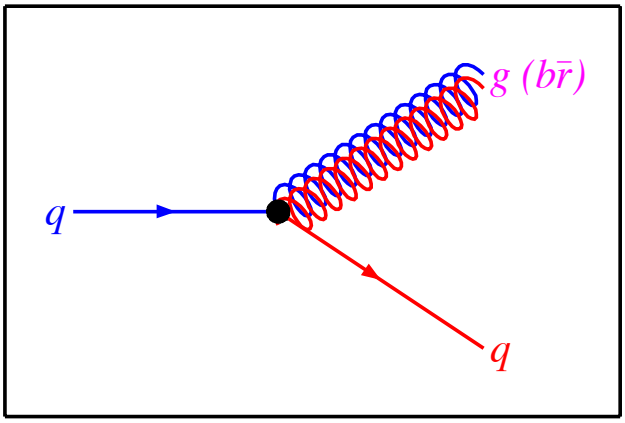


Figure 1.4 The fundamental quark-gluon vertex of QCD.

diagrams whereby gluons are coupling to other gluons. This does not occur in QED as photons do not carry electric charge. Two further consequences arise from this property, making QED and QCD intrinsically different.

Asymptotic freedom refers to the fact that the strength of the strong force (i.e. the coupling strength) increases with increasing distance. This is because in QCD both quark anti-quark loops and gluon-gluon loops contribute to the higher order processes. However, although quark anti-quark loops (like loops in QED) lead to a net reduction in coupling strength with increasing distance the opposite is true for gluon-gluon loops. Because there are more gluons than quarks, the effect from the gluon-gluon loops outweighs that from the quark anti-quark loops and as such gives rise to a net increase in coupling strength with increasing distance. A consequence of asymptotic freedom is that no free quarks or gluons, unlike leptons and photons, are observed in nature and this is why we only observe colourless hadrons in the form of baryons or mesons.

1.2.3 The Electroweak Theory

In analogy to QED and QCD, the quantum field theory of weak interactions is determined by requiring local gauge invariance of the appropriate Lagrangian. The local gauge group of the weak interaction under which the Lagrangian must be invariant is $SU(2)_L$ in the vector space of weak isospin I, where the L subscript refers to the fact that the fermions whose interactions the Lagrangian describes are left-handed (i.e. with intrinsic spin orientated opposite to direction of motion). All left-handed fermions experience the weak interaction and are arranged into pairs or fermion doublets. For the leptons these doublets consist of the same generations shown in Figure 1.2, corresponding to the physical (mass) eigenstates. The weak interaction does not couple to the quark mass eigenstates (i.e. u,s,d) but instead, linear combinations of them (u',s',d') which are determined by the CKM matrix[9].

Flavour changing neutral current reactions are not allowed in the SM. In the original formulation of the SM, neutrinos only experience the weak interaction and so are not predicted to have a right-handed component. In 1968 Glashow, Salam and Weinberg[1],[2],[3] successfully extended the theory of weak interactions to encompass the electromagnetic interaction by using the gauge group $SU(2)_L \otimes U(1)_Y$. Here Y is called *weak hyper-charge* and is related to electric charge Q by $Q = Y + I_3$ where I_3 is the 3^{rd} component of weak isospin. This showed that the electromagnetic and weak forces can be viewed as two components of a single force, called the electroweak force at high energy. In this case preserving local gauge invariance requires four massless fields be introduced. Mixing of these four fields gives the electroweak bosons γ, W^+, W^-, Z. The non-abelian nature of the $SU(2)_L$ group gives rise to self interaction terms and allows W(and Z) bosons to couple to each other. The $U(1)_Y$ group is however abelian and as such the absence of photon-photon couplings in QED is maintained.

1.2.4 Spontaneous Symmetry Breaking and the Higgs Mechanism

A combination of the theories of the electroweak and strong interactions may be done to form a unified theory of all the fundamental forces apart from gravity. The corresponding local gauge group is $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ and the associated Lagrangian, the SM Lagrangian. However as with the individual theories, the fields which must be included in order to preserve local gauge symmetry (these are identified with the γ, W^+, W^-, Z and gluons) are each required to be massless. The same is true for the fermions which under SU(2)L must be massless. Experimentally, however, it has been shown that the W^+, W^-, Z and fermions are indeed massive.

The mechanism of *spontaneous electroweak symmetry breaking* applied to a *non-abelian* theory was introduced by Peter Higgs[10] in 1964, and independently by Robert Brout and Francoise Englert[11], and Gerald Guralnik, C. R. Hagen, and Tom Kibble[12],[13]. It provides a solution to the massless fields. This is what is commonly known as the Higgs mechanism.

The basic idea behind *spontaneous symmetry breaking* is discussed in the following by means of adding to the theory a complex scalar field $\frac{\phi_1 + i\phi_2}{\sqrt{2}}$ with Lagrangian equation 1.1

$$\mathcal{L} = T - V(\phi) = (\partial_\nu \phi)(\partial^\nu \phi) - \frac{1}{2}\mu\phi^*\phi - \frac{1}{4}(\phi^*\phi)^2 \tag{1.1}$$

where $V(\phi)$ is the potential and μ and λ are two free parameters. In order for this Lagrangian to be invariant under global gauge transformations (the associated gauge group is $U(1)$), i.e. symmetric under $\phi \to -\phi$, and there to exist a vacuum state with positive and finite energy, λ must be positive. Imposing $\lambda > 0$, (Figure 1.5) two solutions for μ^2 exist:

1) $\mu^2 > 0$ describes a scalar field with mass μ. ϕ is a self-interacting field with coupling λ. The ground state vacuum expectation value here is $\phi = 0$; satisfying mirror symmetry.

2) $\mu^2 < 0$ describes a case where the mass term has the wrong sign for ϕ since the relative sign between the kinematic and the potential energy is now positive.

The potential has two minima which satisfy equation 1.1. The solution to equation 1.1 are $\phi = 0$ and $\phi = \pm\sqrt{\frac{\mu}{\lambda}}$. The two minima are: $\phi = \pm v$, where $v = \sqrt{\frac{\mu}{\lambda}}$.

Perturbative expansion about the minimum allows to write ϕ like $\phi(x) = v + \eta(x)$, where $\eta(x)$ represents quantum fluctuations and the field has been translated to $\phi = +v$. $\phi = -v$ is easily achieved through mirror symmetry.

Doing the corresponding substitution into the mentioned Lagrangian, we have:

$$\mathcal{L} = \frac{1}{2}(\partial_\nu \eta)(\partial^\nu \eta) - \lambda v^2 \eta^2 - \lambda v \eta^3 - \frac{1}{4}\lambda \eta^4 + const. \tag{1.2}$$

Now the mass term of the field η has a correct sing,

$$m_\eta = \sqrt{2\lambda v^2} = \sqrt{-2\mu^2}, \lambda = -1. \tag{1.3}$$

The higher order terms (η^3 and η^4) correspond to the self-interactions in η, similar to ϕ^4 in 1.1. This new Lagrangian gives us an accurate picture of physics for scalar particles. It now has a mass term because of the way it was generated, i.e. ϕ was expand as a function of η around the $\phi = +v$.

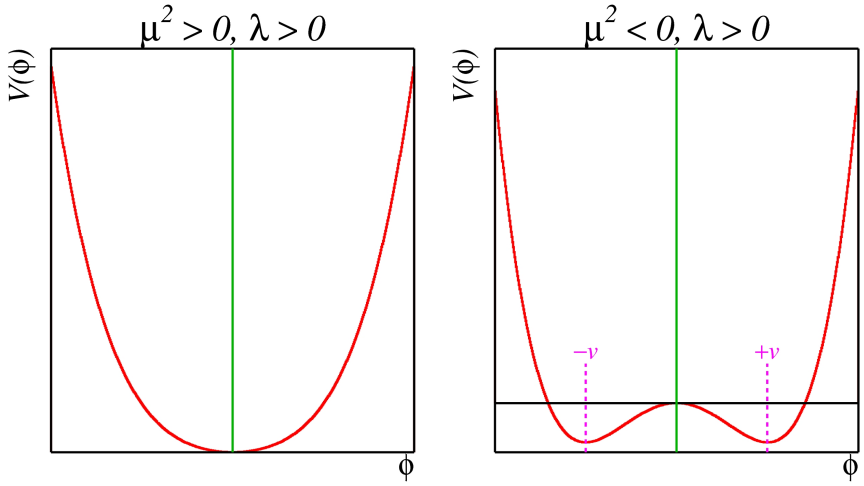


Figure 1.5 The potential $V(\phi) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4$ for $\mu^2 > 0$, and for $\mu^2 < 0$ where $\lambda > 0$ in both cases.

Spontaneous symmetry breaking corresponds to the choice in selecting the ground state $\phi = +v$. In physics, spontaneous symmetry breaking occurs when a system belonging to a particular symmetry group goes into a vacuum state that is not

symmetric.

Taking into account the previous statement, a look into the spontaneous breaking of a global symmetry is going to be useful in order to describe a Lagrangian for which the mass of gauge bosons can be generated. Repeating the above procedure for a complex scalar field $\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)$, which is invariant under the phase transformation $\phi \to e^{i\alpha}\phi$, the Lagrangian is

$$\mathcal{L} = (\partial_\nu \phi)^*(\partial^\nu \phi) - \mu^2 \phi^* \phi - \lambda(\phi^* \phi) \tag{1.4}$$

By substituting the field $\phi = \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)$ into equation 1.4, becomes

$$\mathcal{L} = \frac{1}{2}(\partial_\nu \phi_1)^2 + \frac{1}{2}(\partial_\nu \phi_2)^2 + \frac{1}{2}\mu^2(\phi_1^2 + \phi_2^2)^2 \tag{1.5}$$

After minimizing the potential, we can look at the cases $\lambda > 0$, $\mu^2 < 0$. The minima for $v(\phi)$ exist in a plane ϕ_1, ϕ_2, with radius v such that $v^2 = \phi_1^2 + \phi_2^2$, where $v^2 = -\frac{\mu^2}{\lambda}$, that translates ϕ to a minimum energy position.

In a simple case, lets set $\phi_1 = v$ and $\phi_2 = 0$, we can expand the Lagrangian about the vacuum in terms of the fields η and ϵ

$$\phi(x) = \sqrt{\frac{1}{2}(v + \eta(x) + i\epsilon(x))} \tag{1.6}$$

that if we included it into the equation 1.4, becomes $\mathcal{L} = \frac{1}{2}(\partial_\nu \epsilon)^2 + \frac{1}{2}(\partial_\nu$ $eta)^2 + \mu^2\eta^2 +$ *constant + higher orders*, where the third term has the form of a mass term for η with a mass of $m_\eta = \sqrt{-2\mu^2}$. As for ϵ there is not apparent mass term, the theory classifies it as a massless scalar Goldstone boson[14, 15, 16], that are bosons that appear in modes which exhibit spontaneous symmetry breaking.

At this point, lets move to look at the spontaneous breaking of local gauge symmetry, starting with the request that the Lagrangian must be invariant under $U(1)$ gauge transformations in ϕ. This is done by transforming $\phi \to e^{i\alpha(x)}\phi$ and replacing ∂_μ with a covariant derivative, $D_\mu = \partial_\mu - ieA_\mu$. The gauge field A_μ transforms as $A_\mu \to A_\mu + \frac{1}{e}\partial_\mu \alpha$. Where A_μ couples to the Dirac particle charge $-e$. Then, the Lagrangian is written as

$$\mathcal{L} = (\partial_\mu - ieA_\mu)\phi^*(\partial_\mu - ieA_\mu)\phi - \mu^2\phi^*\phi - \lambda(\phi^*\phi)^2 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{1.7}$$

plus interaction terms, where $v = \pm\sqrt{-\frac{\mu^2}{\lambda}}$.

Now the situation is: a massless Goldstone boson ϵ, a massive scalar η, and a massive vector A_μ with their masses: $m_\epsilon = 0$, $m_\eta = \sqrt{2\lambda\nu^2}$, $m_A = ev$.

By having generated a mass for the gauge field, we still need to solve de problem relative to the generation of a massless Goldstone boson. Giving mass to A_μ, the

number of degrees of freedom is taking from two to three, requiring that we deduce that all the fields must not correspond to distinct particles.

Approximating ϕ to the lowest order in ϵ, i.e. $\phi = \sqrt{\frac{1}{2}}(v+\eta+i\epsilon) \approx \sqrt{\frac{1}{2}}(v+\eta)e^{i\frac{\epsilon}{v}}$. And substituting different fields h, θ, A_μ, where h is real:

$$\phi \to \sqrt{\frac{1}{2}}(v+h(x))e^{\frac{i\theta(x)}{v}}; \tag{1.8}$$

$$A_\mu \to A_{mu} + \frac{1}{ev}\partial_\mu\theta \tag{1.9}$$

that gives to the Lagrangian the following form:

$$\mathcal{L} = \frac{1}{2}(\partial_{mu}h)^2 - \lambda v^2h^2 + \frac{1}{2}e^2v^2A_\mu^2 - \lambda vh^3 - \frac{1}{4}\lambda h^4 + \frac{1}{2}e^2A_\mu^2h^2 + ve^2A_\mu^2h - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \tag{1.10}$$

It is possible to see that the Goldstone boson does not appear in the equation and the extra degree of freedom corresponds to the ability to make gauge transformations. The Lagrangian describes two interactive massive particles A_μ and h. A_μ is a vector gauge boson and h is a massive scalar known as the Higgs particle. With this mathematical computation, the Goldstone boson now represents a longitudinal polarization for the massive gauge boson, this is known as the Higgs Mechanism.

1.3 Higgs boson mass constrains

In the Standard Model (SM) there are four types of gauge vector bosons (W, Z, photon(γ) and gluon) and twelve types of fermions (six quarks and six leptons)[1, 2, 3, 4]. These particles have been observed experimentally. At present, all the data obtained from the many experiments in particle physics are in agreement with the SM. We already know that the Higgs boson, according to the SM, is the responsible for giving masses to all the particles. In this sense the Higgs particle occupies a unique position. But the theory itself does not predict the Higgs boson mass, so, in order to understand how the Higgs's search have been perform during the last decades, lets make a review of the theoretical and experimental constrains to such a value, including the most recent LHC's results, where the discovery of a new Higgs-like boson have boosted this already exciting research.

1.3.1 Theoretical Constrains

As mentioned earlier, the Higgs boson mass is not predicted by the SM, nevertheless there are several theoretical ways to define boundaries to the Higgs mass value. Such a constrains have been classified as follow:

Unitarity

Define as the requirement that the total scattering probability for a process, obtained from integrating over all contributing Feynman diagrams, remains less that 1.The theory of electroweak interaction of Fermi violates unitarity at the electroweak scale $\sqrt{s} \approx G_{\mu}^{-\frac{1}{2}}$ because it assumes point-like interactions. The introduction of massive intermediate bosons resolves the problem for low energy. However, certain processes involving the longitudinal components of the vector bosons are expected to violate unitarity at tree-level. An example of such a process is $W^+W^- \to W^+W^-$ longitudinal scattering (see Figure 1.6) which gets contributions from Z and γ, leading to its cross section increasing at high energy in proportion to the square of the centre of mass energy or $\sqrt{s}$[28]. Unitarity can be restored by adding Higgs exchange diagrams, establishing an upper bound on the Higgs mass of $2v\sqrt{\pi} \approx 800$ GeV.

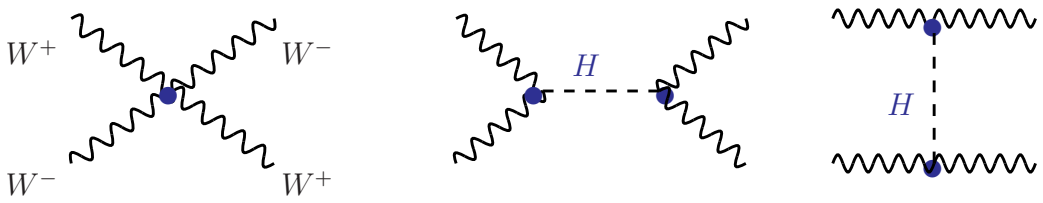


Figure 1.6 Some Feynman dyagramm that contribute to the $W^+W^- \to W^+W^-$ scattering.

Triviality

Referring to section 1.2.4, the mass of the Higgs boson is given by $m_h = \sqrt{-2\mu^2} = v\sqrt{2\lambda}$. This represents the leading order expression and will be modified once higher order corrections are accounted for, examples of which are shown in Figure 1.7. These corrections give rise to divergences which can be accounted for using a renormalisation procedure. After applying this renormalisation it becomes evident that the Higgs self-coupling λ diverges with increasing energy scale. Assuming the Higgs self-coupling is larger than the top quark Yukawa coupling, it varies with energy like $\lambda^2 ln(Q^2)$. Since it is assumed that the SM is valid at all energies, λ must be zero. This implies that the SM is valid up to a cut-off energy scale (Λ) at which new physics will begin to appear. In order that it produces meaningful predictions at energies below this cut-off the perturbativity of the SM theory must be maintained and as such the Higgs self coupling must remain finite. Since the Higgs coupling is proportional to the square of the Higgs mass, an *upper limit* on its value is predicted, depending on Λ. This is called the triviality bound on the mass of the Higgs boson. Like an example, a cut-off energy of 10 TeV the upper limit imposed by this constraint gives a $m_h \approx 500$ GeV.

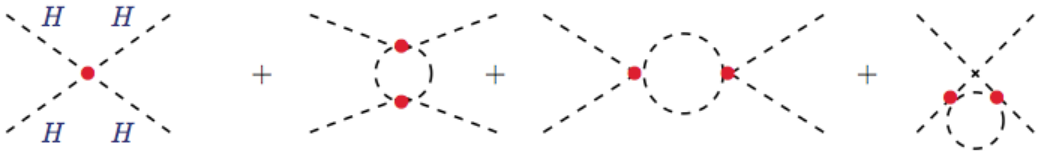


Figure 1.7 Feynmann dyagramms for the three level of the self-intereracion at one *loop* of the Higgs boson.

Vacuum stability

A *lower constraint* on the Higgs boson mass is derived by assuming the top quark Yukawa coupling is larger than the Higgs self-coupling. This implies low Higgs masses $-m_h < m_t-$ where the coupling to top and weak bosons becomes large. In this scenario, the Higgs potential will develop a global minimum at large energy scales, thereby creating an unstable vacuum and preventing spontaneous symmetry breaking. Imposing the same cut-off energy scale Λ at which the SM is valid to, a lower constraint on the Higgs boson mass is imposed in order to maintain vacuum stability. For a cut-off energy scale of the electroweak scale ($\Lambda = 10^3$) $m_h \approx 70$ GeV.

The combined effect of the triviality and vacuum stability requirements is shown in Figure 1.8, showing the allowed Higgs mass window as a function of the cut-off energy scale Λ[17], where the bands represent the upper and lower limit theoretical uncertainties and enclose the allowed Higgs boson mass. The limits were derived assuming a top mass $m_t = 175 \pm 6$ GeV and strong coupling constant $\alpha_s = 0.118 \pm 0.002$.

1.3.2 Experimental Constrains: *including the LHC's July 4^{th} 2012 discovery*

Experimental constrains on the SM Higgs mass have been established in two ways, in form of limits coming from direct Higgs searches performed at colliders like LEP at CERN and TeVatron at Fermilab, and indirect limits, arising from precision measurements of the electroweak parameters.

The LEP machine was an e^+e^- collider which was operative at CERN from 1989 to 2000. In the first phase of its operations (LEP I) it provided collision at $89 < \sqrt{s} < 93$ GeV to perform precision studies on the recently discovered Z boson, while in the second phase (LEP II) the search for the Higgs boson became one of its main goals, and collision where recorded at increasing energy up to $\sqrt{s} = 210$ GeV. The LEP machine provided data to four detector experiments: ALEPH [75], DELPHI [76], L3 [77] and OPAL [78].

The main Higgs production mechanism at LEP was the Higgs-strahlung processes, in which an Higgs boson is radiated by a virtual Z boson: $e^+e^- \to Z^* \to ZH$. And all the possible detectable decay modes of H and Z have been used in the search.

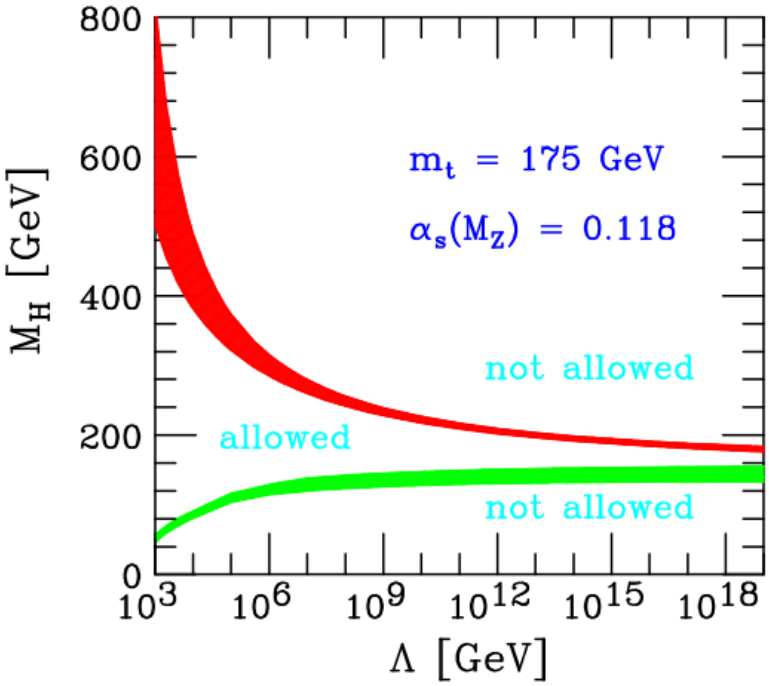


Figure 1.8 Theoretical upper and lower limits on the mass of the Higgs boson as a function of cut-off energy scale Λ. The upper limits are provided by the triviality bound and the lower limits by the vacuum stability bound.

While some initial hints of a Higgs signal with mass around 115 GeV was seen, in the final combined result of the search for the Higgs boson performed by the four experiments didn't show any relevant excess, and the final result is shown in Figure 1.9: the test statistics is $-2\ln Q = -2\ln\frac{\mathcal{L}_f}{\mathcal{L}_\emptyset}$, where $\mathcal{L}_\emptyset$ and $\mathcal{L}_f$ are the likelihood of the background only and signal plus background hypotheses respectively.

From Figure 1.9 one can deduce that up to a Higgs mass of 114.4 GeV the observed data are consistent with the background only hypothesis.

The TeVatron is a proton-antiproton collider operating in the so called RUN II at a center of mass energy of 1.96 TeV and it has been taking data up to 2011 providing data to two detector experiments: CDF [79] and D0 [80].

The main Higgs production mechanism at the TeVatron collider was the associate production including also the W boson ($p\bar{p} \to VH, V = W^{\pm}, Z$), while the main decay channels include also the decay to pairs of vector bosons ($H \to ZZ^*$ and $H \to W^+W^-$) because of the wider mass range accessible at the TeVatron. The results of the combined search of CDF and D0 are shown in Figure 1.10, [81]: the 95% confidence level upper limit on the ratio of the Higgs boson production to the SM expectation is shown as a function of the Higgs boson mass.

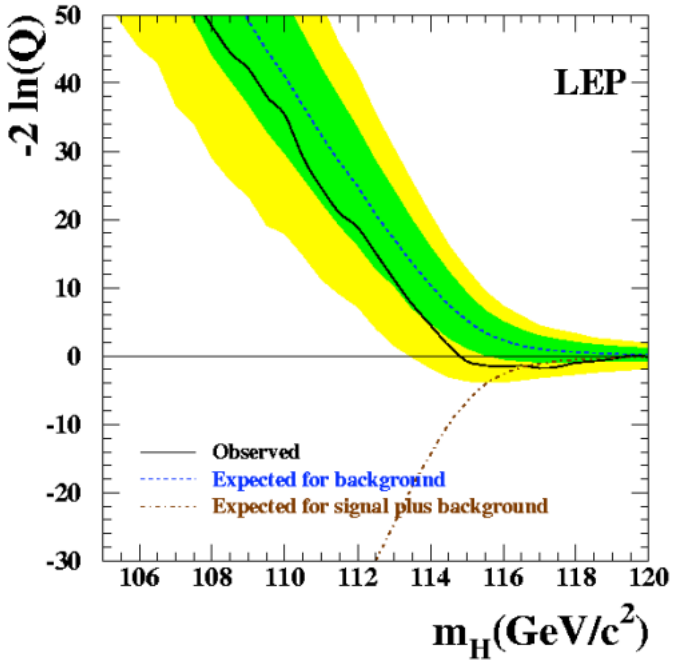


Figure 1.9 Combined results of the direct Higgs search preformed by the four experiments at LEP.

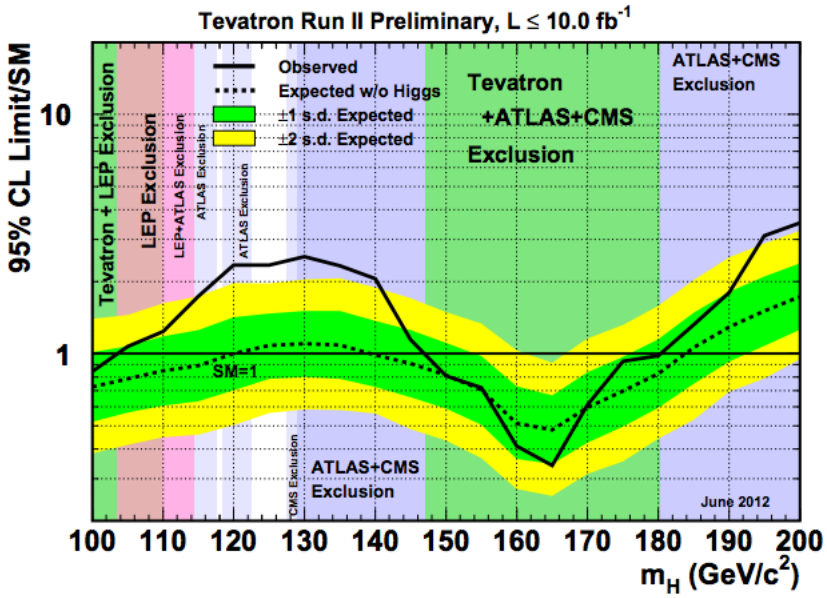


Figure 1.10 Combined results of the Higgs searches by the CDF and D0 collaborations.

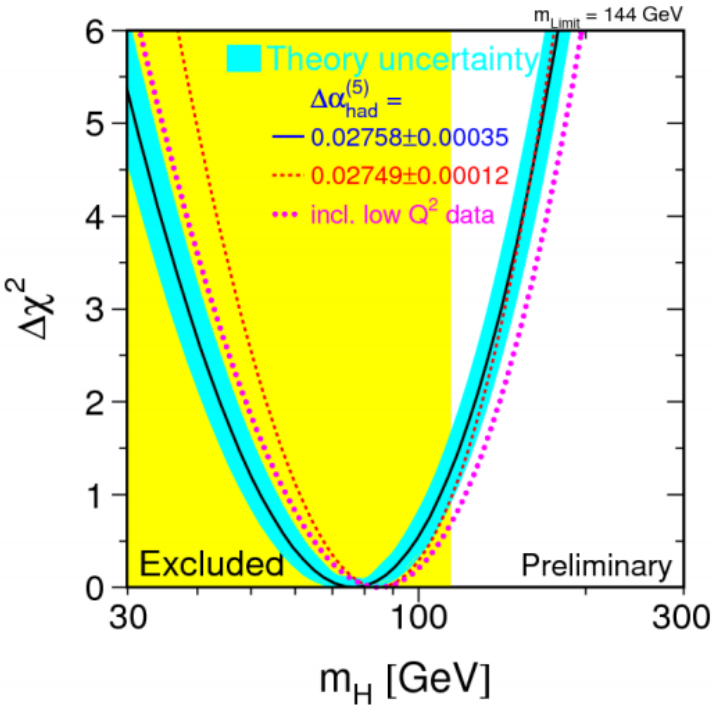


Figure 1.11 Variation of the χ^2 of the electroweak fit as a function of the Higgs boson mass

As can be seen the observed limit goes below unity in the interval $147 < m_H < 180$ GeV and therefore the presence of the Higgs boson is excluded in this mass range with a 95% confidence level.

With only Higgs boson as the missing piece of the SM predictions, it is obvious to attend to predict its mass by fitting all data within the SM framework, having it as one of the free parameters. The variations of the χ^2 of this fit to the data collected by the LEP, TeVatron and SLC accelerators are shown in Figure 1.11. The main result of this fit is that the low mass region (compatible with LEP and TeVatron results) is favoured, but also the high-mass region is not excluded.

Higgs Search, and discovery of a Higgs-like particle at LHC

There are essentially four mechanics for the single production of the SM Higgs boson at hadron collider[18]; some Feynman diagrams are show in the Figure 1.12.

The total cross sections, are displayed in Figure 1.13 for the LHC at two different center of mass energies: $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV as a function of the Higgs mass. Lets summarised below the main features of each production channel:

$gg \to H$

This is the dominant production process at the LHC, up to masses $M_H \approx 1$ TeV. The most promising detection channels are[19] $H \to \gamma\gamma$ for $M_H \leq 130$ GeV and slightly above this mas value, $H \to ZZ^* \to 4l^{\pm}$ and $H \to WW^* \to ll\nu\nu$ with $l = e, \mu$, for masses below $2M_W$ and $2M_Z$ respectively. For higher masses, $M_H > 2M_Z$, it is the golden channel $H \to ZZ \to 4l^{\pm}$, witch from $M_H > 500$ GeV can be complemented by $H \to ZZ \to \nu\nu l^+ l^-$, $H \to WW \to 4l^{\pm}$ and −our dissertation analysis− $H \to ZZ \to l^+ l^- q\bar{q}$ to increase statistics.

$q\bar{q} \to HV$

The associated production with gauge bosons, with $H \to b\bar{b}$ and possibly $H \to WW^* \to l^+\nu jj$, is the most relevant mechanism at TeVatron[20] [$gg \to H \to WW \to l\nu l\nu$ being important for Higgs masses close to 160 GeV]. At the LHC, this process plays only a marginal role; however, the channels $HW \to l\nu\gamma\gamma$ and eventually $l\nu b\bar{b}$ could be useful for the measurement of Higgs couplings.

WW/ZZ fusion

This process has de second largest cross section at the LHC. For several reasons, the interest in this process has grown: it has a large enough cross section an one can use cuts, forward-jet tagging, mini-jet veto for low luminosity as well as triggering on the central Higgs decay products[21] which render the background comparable to the signal, therefore allowing precision measurements.

$pp \to t\bar{t}H$

Higgs boson production in association with top quarks, with $H \to \gamma\gamma$ or $b\bar{b}$, can be observed at the LHC and direct measurement of the top Yukawa coupling, as well as an unambiguous determination of the CP violation of the Higgs boson can be possible, in spite that $pp \to t\bar{t}H \to t\bar{t}b\bar{b}$ may be subject to a too large jet background[22].

The discovery of this new particle has been made public with a press conference held at CERN on July 4th 2012, and it is based on the analysis and the combination of both the 2011 and 2012 datasets, studied independently by the two collaborations (ATLAS and CMS).

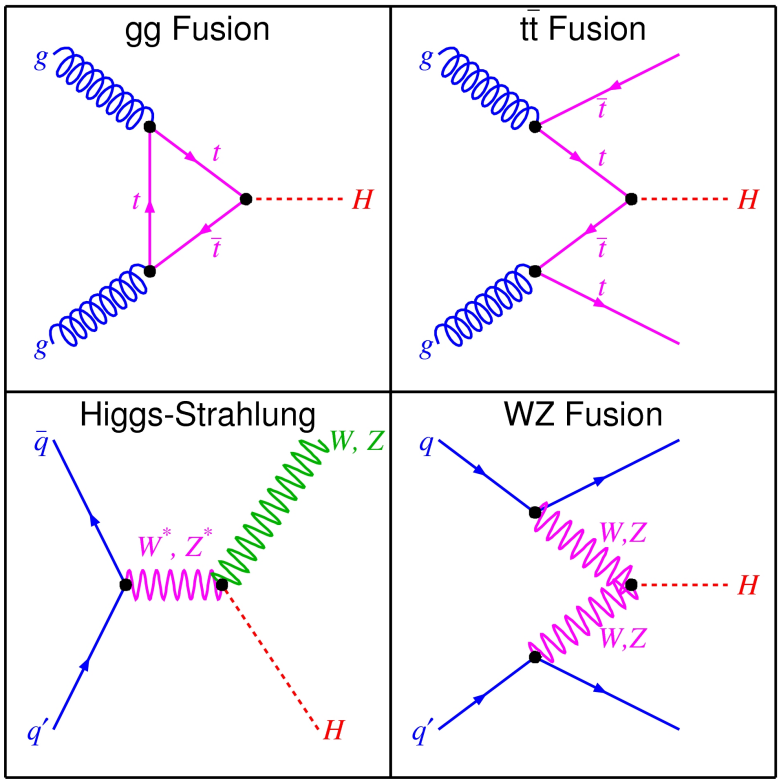
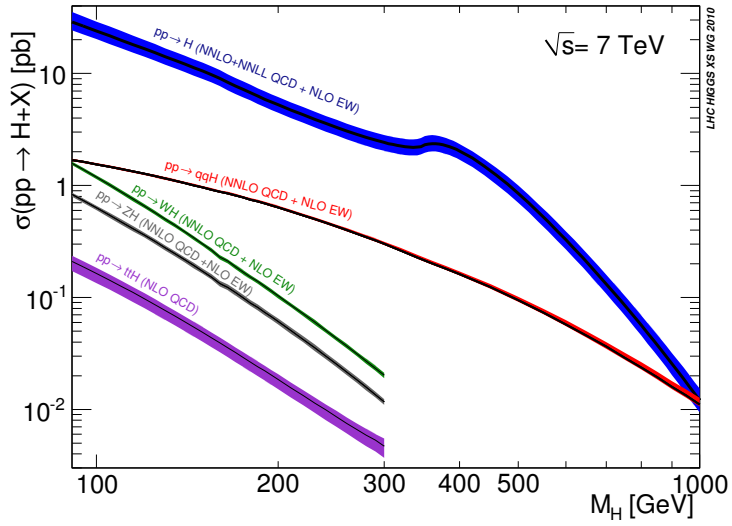
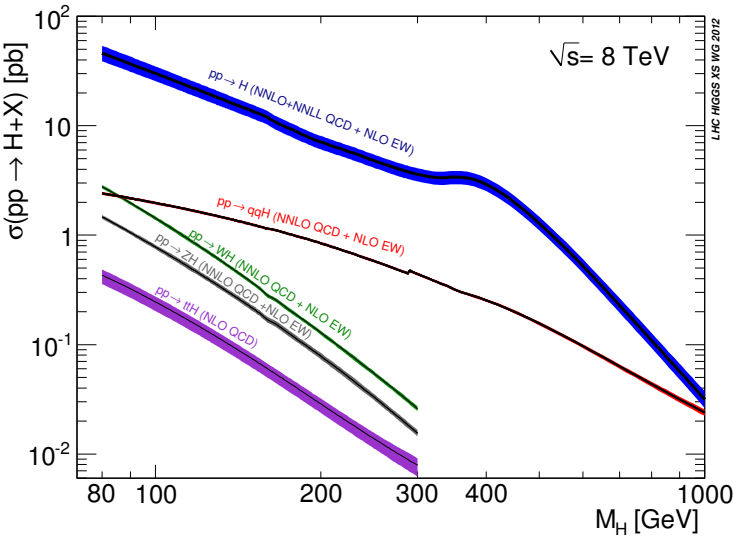


Figure 1.12 The most important processes for Higgs production at hadron colliders. Gluon fusion, vector boson fusion, Associative production with W and an example of the diagrams having associative production with a top pair.



(a) Integrated luminosity recorded in 2011



(b) Peak luminosity reached in 2011

Figure 1.13 Standard Model Higgs boson production cross sections at (a) $\sqrt{s} = 7$ TeV and (b) $\sqrt{s} = 8$ TeV. Transition for VBF at $MH = 300$ GeV at 8 TeV is due to change from ZWA to complex-pole-scheme.

Figure 1.14 shows the evolution of the p-value [111] of the combined Higgs search in the ATLAS experiment as a function of time (i.e. available integrated luminosity): the dashed lines stand for the expected p-value distributions while the solid lines represent the observed ones.

The p-value is basically the probability that, given a certain hypothesis (e.g. the background-only hypothesis), the data have fluctuations greater than the observed ones. The aim of such a measurement is to exclude at 5σ the background-only hypothesis.

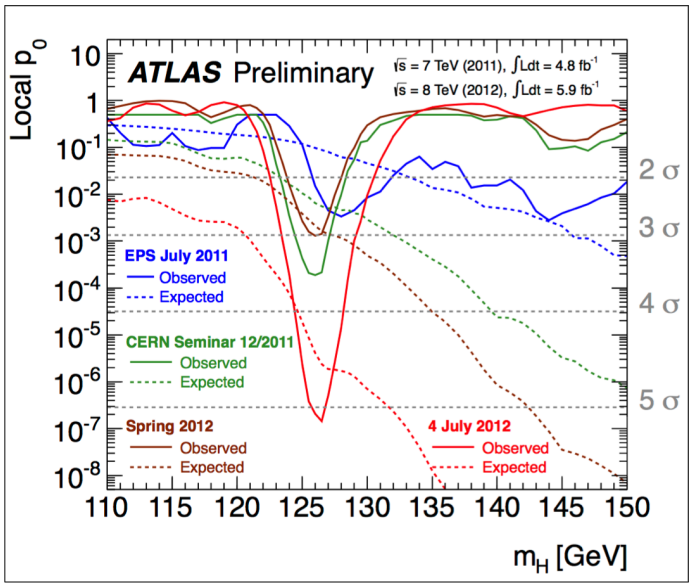


Figure 1.14 Evolution of the p-value measured in the combined Higgs boson search by the ATLAS experiment as a function of time. 2012 results include the full 2011 data sample combined with the amount of data collected in 2012 (at $\sqrt{s} = 8$ TeV) up to the closure for the reference public document.

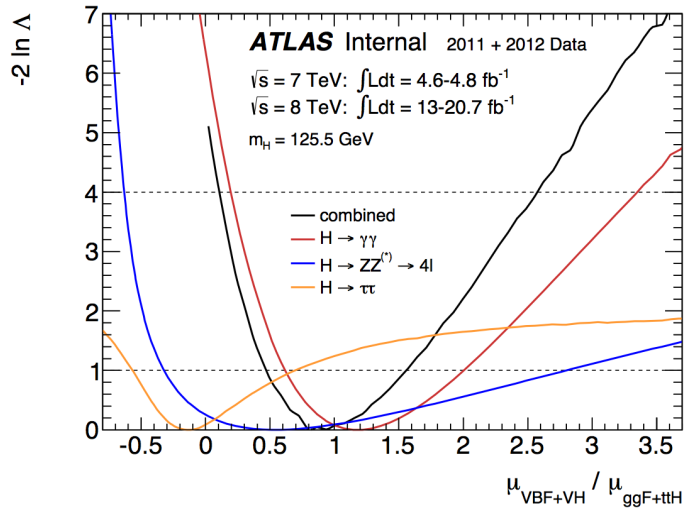


Figure 1.15 Likelihood curves for the ratio $\mu VBF+VH/\mu ggF+t\bar{t}H$ from the $H \to \gamma\gamma$, $H \to ZZ^* \to 4l$ and $H \to \tau^+\tau^-$ channels and their combination for a Higgs boson mass hypothesis of $m_H = 125.5$ GeV. The branching ratios and possible non-SM effects coming from the branching ratios cancel in $\mu VBF+VH/\mu ggF+t\bar{t}H$, hence the different measurements from all three channels can be compared and combined.

Recently, (March, 2013) such a observation of a new particle in the search for the Standard Model (SM) Higgs boson at the LHC, reported by the ATLAS[23] and CMS[24] Collaborations, has been update into a combined signal strength value for low mass resolution channels $H \to WW^* \to l\nu l\nu$, $H \to \tau^+\tau^-$ and $H \to b\bar{b}$[25]. The mass, signal strength and couplings measurements have been updated using up to $4.8 fb^{-1}$ of pp collision data at $\sqrt{s} = 7$ TeV and about $21 fb^{-1}$ at $\sqrt{s} = 8$ TeV for the two channels $H \to \gamma\gamma$[26] $H \to ZZ^* \to 4l$[27]. The Figure 1.15 shows the Likelihood curves for the ratio different channels and their combination for a Higgs boson mass hypothesis of $m_H = 125.5$ GeV.

1.4 Beyond the Standard Model

Despite being able to explain with high precision most of the experimental data that has been produced until now, the SM suffers from several theoretical problems:

- No dark matter candidate is provided by the SM.
- It does not explain the gravitational interaction.
- The level of CP violation is not sufficient to explain the matter anti-matter asymmetry seen in the universe.
- It does not explain the hierarchy problem, i.e. why gravity is so weak compared to the other interactions.
- Fine tuning is required to deal with divergences in the Higgs sector.

Because of these reasons and indeed others not discussed, it is a widely held opinion within the scientific community that the SM is an effective theory which we currently probe at low energy. The general theory will begin to become accessible when the predictions of the SM start to become incorrect. More precise determination of the free parameters of the SM will allow the scale at which this happens to be better understood.

Several beyond the SM theories exist which describe the SM predictions at low energy. Perhaps one of the most popular is supersymmetry which suggests a new symmetry between fermions and bosons and provides solutions to a number of the problems associated with SM. For example its models commonly provide a dark matter candidate and it also provides a solution to divergences in the Higgs sector.

1.4.1 Two-Higgs-Doublet models(2HDM)

In this form the SM, the Higgs mechanism constitutes only a minimal configuration to implement the breaking of the electroweak symmetry and the generation of particle masses. A simple extension of the SM Higgs sector is given by the addition of a second complex Higgs doublet [152], giving rise to five Higgs bosons:

Two CP-even scalar fields h and H, one pseudoscalar A (CP-odd), and two charged fields $H^{\pm}$. These Two-Higgs-Doublet models (2HDM) are phenomenologically interesting since they can explain the generation of the baryon asymmetry in the universe [8] and are an important ingredient of axiom models that are designed to explain the dark matter content of the universe [153]. Finally, the minimal supersymmetric SM [154] contains two Higgs doublets as well. Four different types of 2HDMs can be distinguished, depending on the different coupling of the two scalar fields h and H to fermions and weak gauge bosons. In type-I models all quarks couple to just one of the Higgs doublets, while in type-II models the right-handed up-type quarks couple to one Higgs doublet and the right-handed down-type quarks

to the other doublet. Type-III and type-IV models differ only from type-I and type-II models in their couplings to the leptons.

Recent detailed reviews on 2HDMs can be found in Refs. [155, 156]. Since the discovery of the new boson at the LHC, 2HDMs have attracted much attention in phenomenological studies [157, 158, 159, 160, 161], which provide a strong incentive for dedicated experimental investigations in this direction. Searches for generic 2HDMs have been performed by the CDF collaboration at the TeVatron[162, 163]. The rate of the Higgs-like boson at 125 GeV in the two-photon channel provides also constraints on 2HDMs[164], mainly reducing the parameter space of type-II models. The analysis presented in this note investigates the possibility that the boson observed by the ATLAS and CMS experiments at a mass of 125 GeV originates from a Higgs boson that is part of a 2HDM. In particular, it is assumed that the observed particle is the low mass Higgs h of the 2HDM.

Chapter 2

CERN, the LHC and the ATLAS Experiment

Founded in 1954, the CERN laboratory sits astride the Franco-Swiss border near Geneva. It was one of Europe's first joint ventures and now has 20 member states. The name CERN is derived from the acronym for the French *Conseil Européen pour la Recherche Nucléaire*, or European Council for Nuclear Research, a provisional body founded in 1952 with the mandate of establishing a world-class fundamental physics research organization in Europe. At that time, pure physics research concentrated on understanding the inside of the atom, hence the word "nuclear". At this moment, the understanding of matter goes much deeper than the nucleus, and the main area of research at CERN is particle physics − the study of the fundamental constituents of matter and the forces acting between them.

At CERN, the European Organization for Nuclear Research, physicists and engineers are probing the fundamental structure of the universe. They use the world's largest and most complex particle accelerator to study the basic constituents of matter or the fundamental particles. The particles are made to collide together at close to the speed of light. The process gives the physicists clues about how the particles interact, and provides insights into the fundamental laws of nature.

The instruments used at CERN are purpose-built particle accelerators and detectors. Accelerators boost beams of particles to high energies before the beams are made to collide with each other. Detectors −like ATLAS− observe and record the results of these collisions.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC), conceptualised around a quarter of century back, is built in a circular tunnel 27 km in circumference. The tunnel is buried around 50 m to 175 m underground. It located between the Swiss and French borders on the outskirts of Geneva.

The first beams were circulated successfully on 10th September 2008. Unfortunately on 19^{th} September a serious fault developed damaging a number of superconducting magnets. The repair required a long technical intervention. The LHC beam did not see beam again before November 2009. First collisions took place on 30^{th} March 2010 with the rest of the year mainly devoted to commissioning. 2011 was the first production year with over $5fb^{-1}$ delivered to both ATLAS and CMS. The 2012 started with over $6fb^{-1}$ delivered by the time of the summer conferences. Data that allowed for the announcement of the discovery of a Higgs-like particle on 4^{th} July 2012 mentioned above.

It is a proton-proton (pp) collider, and the collision were delivered at $\sqrt{s} = 7$ TeV in 2010 and 2011, while they are being collected at $\sqrt{s} = 8$ TeV during 2012. One of the crucial parameters for the discovery power of a particle collider is the *instantaneous luminosity* since it is proportional to the event rate $\frac{dN}{dt}$:

$$\frac{dN}{dt} = \mathcal{L} \times \sigma \tag{2.1}$$

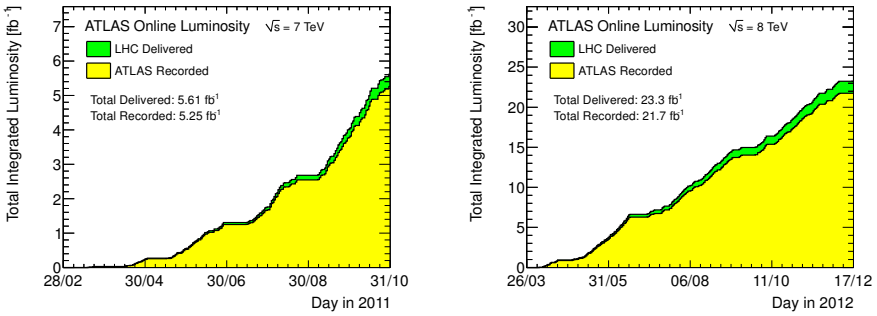
Where σ is the cross section of the considered process. The instantaneous luminosity of a particle accelerator depends on its intrinsic features:

$$\mathcal{L} = \frac{N_p^2 f k}{4\pi R^2} \tag{2.2}$$

where N_p is the number of protons in each bunch, f is the revolution frequency of the protons in the accelerating ring, k is the number of bunches circulating in the beam and R is the mean radius of the proton distribution on the plane orthogonal to the beam direction.

The instantaneous luminosity delivered by the LHC in 2011 reached the value of 3.65×10^{33} $cm^{-2}s^{-1}$ at its maximum, where the design peak luminosity was 10^{34} $cm^{-2}s^{-1}$. This high luminosity is reached with 1380 (2808 from the design) bunches per beam, each of them containing 10^{11} protons. The bunches have very small transverse spread, about 15μm in the transverse direction, and the longitudinal length is about 7 cm. In the design of the LHC the bunches should have crossed every 25 ns, giving a collision rate of 40 MHz, while the actual bunch spacing reached in 2011 and 2012 is of 50 ns. These parameters achieved in 2011 and 2012 allowed an integrated luminosity showed in Figure 2.1.

Part of the acceleration chain and the different positions of the LHC's experiments are showed in Figure 2.2: after their production and an of 1.4 GeV, the Super Proton Synchrotron raises their energy up to 450 GeV before injecting them into



(a) Integrated luminosity recorded in 2011 (b) Integrated luminosity recorded in 2012

Figure 2.1 The integrated luminosity as a function of time delivered by LHC (green) and recorded by ATLAS (yellow) in 2011 and 2012.

Feature	design value	actual value
beam energy [TeV]	7	4
bunch spacing $[ns]$	25	50
peak luminosity $[cm^{-2}s^{-1}]$	10^{34}	8×10^{33}
mean number of interaction per bunch crossing	23	20
number of bunches	2808	1380
protons per bunch	1.15×10^{11}	1.67×10^{11}
bunch transverse dimensions $[\mu m]$	15	~ 30

Table 2.1 Main features of the LHC. The first column contains the values as in the LHC design, the second column contains the actual value of the features. The actual features include both 2011 and 2012 runs

the LHC. Ones there, the protons are accelerated in the two opposite directions up to the colliding energy of 3.5 TeV (2011) or 4 TeV (2012) per beam.

Since LHC accelerates two beams of same sign particles, two separate accelerating cavities and two different magnetic fields are needed: LHC is equipped with 1232 superconducting magnets and 16 radiofrequency cavities which bend and accelerate the proton beams in the two parallel beam lines in the machine. The magnetic field used to bend such energetic proton beams is of 8.3 T and to reach such a magnetic fields the superconducting magnets are cooled down to 1.9 K and a 13 kA current circulates inside them.

The LHC provides collisions in four collision points along its circumference where detector experiments located: ALICE (*A Large Ion ColliderExperiment*), ATLAS (*A Toroidal Lhc ApparatuS*), CMS (*Compact Muon Solenoid*) and LHCb (*Large Hadron Collider beauty*). ATLAS and CMS are multi-purpose detectors, while ALICE and LHCb are focused on more specific studies: (See Figure 2.2) ALICE focuses on the quark-gluon plasma produced in heavy-ions collisions[1], while LHCb focuses on the study of CP violation processes occurring in b and c hadron decays.

[1]The LHC is able to accelerate and collide lead ions at $\sqrt{s} = 2.76$ TeV per nucleon, and ions collisions are foreseen each year in the LHC program. Not part of our actual studies.

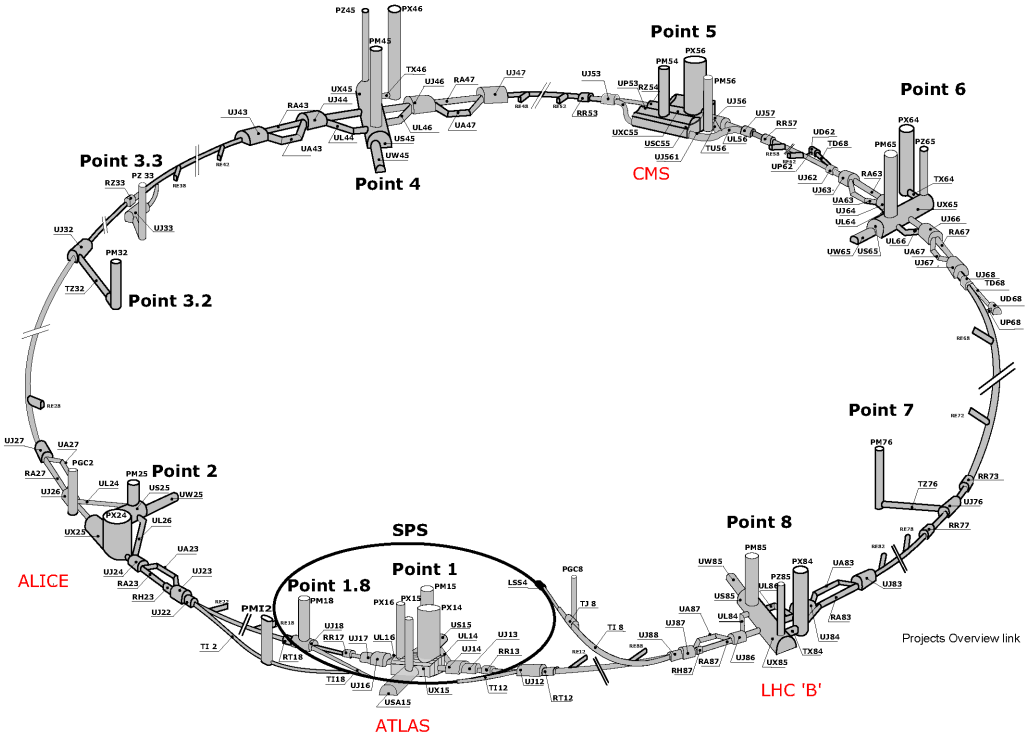


Figure 2.2 The LHC particles accelerator, in which it is possible to see the SPS and the different beam's collision points with their corresponding experiment.

2.2 The ATLAS Detector

The ATLAS detector is one of the four main experiments recording the collisions provided by the LHC. It is 20 m tall and 45 m long and weights more than 7000 $tons$.

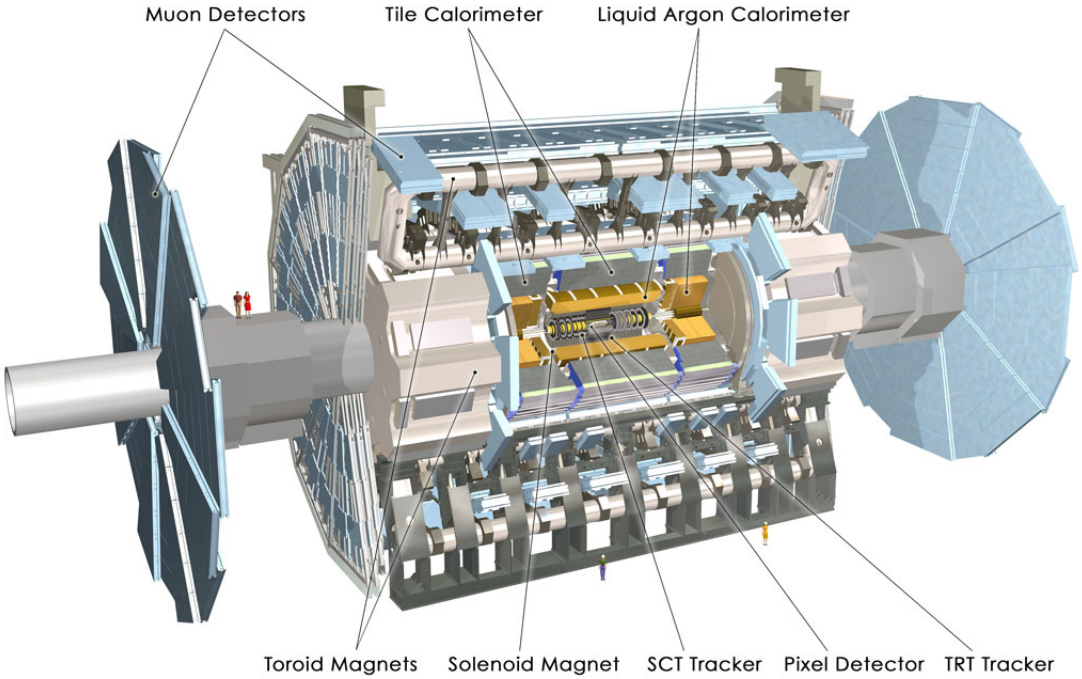


Figure 2.3 The ATLAS Detector: all the subdetectors it is composed of are shown.

The structure has a cylindrical shape centred at the interaction point with its axis along the beam line, and it is composed of several concentric subdeterctors which measure different features of the particles generated in the pp collision as they fly from the center of the detector to the outer part, as shown in Figure 2.4. From the innermost to the outermost layer, the ATLAS experiment is composed of (see Figure 2.3):

- An inner tracking system to detect charged particles and measure their momentum and direction.
- A solenoidal superconducting magnet providing a uniform magnetic field along the beam axis in which the inner detector is immersed.
- An electromagnetic calorimeter to measure the energy deposited by electrons and photons.
- An hadronic calorimeter to measure the energy deposited by hadrons.
- A muon spectrometer, that is a tracking system for the measurement of muons as they travel throughout all the detector and are the only particles reaching the outer part.

- An air-cored superconducting toroidal magnet system which provide the magnetic field to the muon spectrometer.

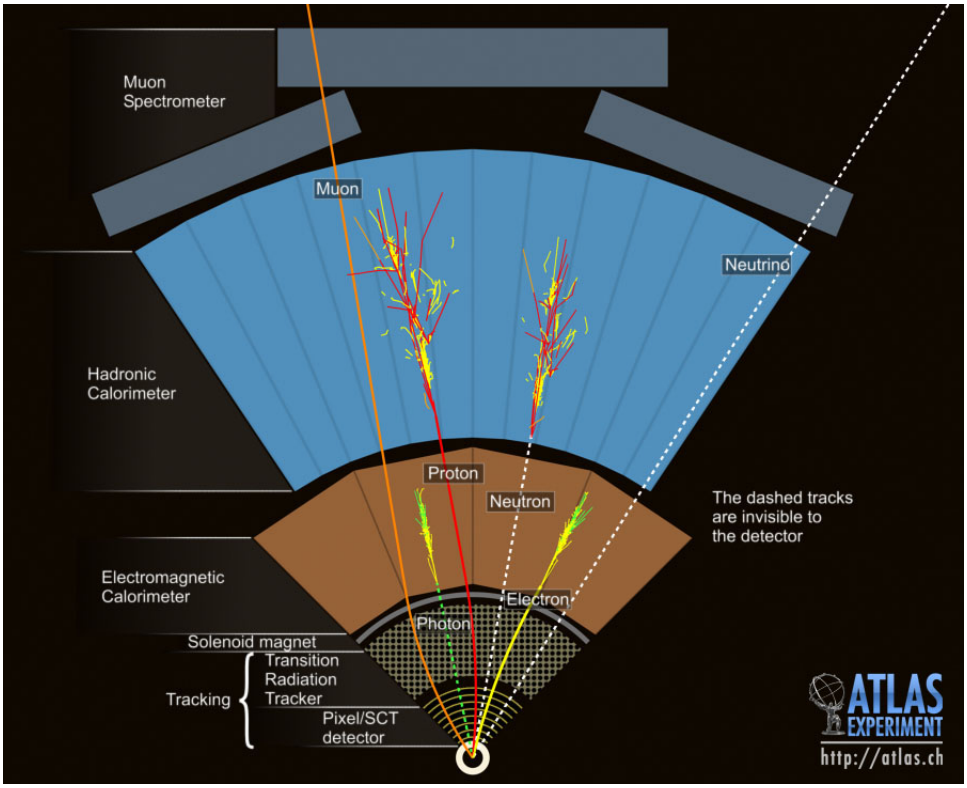


Figure 2.4 Schema of the detection of the particles produced in a proton collision while they travel through the several layers of the ATLAS detector.

In the following sections details about the structure of the subdetectors are be given, as well as some insight about how they work and their performances.

2.2.1 ATLAS Coordinate System

The ATLAS coordinate system is a cartesian right-handed coordinate system, with the nominal collision point as the origin (Figure 2.5), the z axis is along the beam line and the $x - y$ plane is the plane perpendicular to the beam line.

The x axis points to the center of the LHC ring, while the y axis goes upwards. The azimuthal angle ϕ is defined around the beam axis, while the polar angle θ is the angle from the z axis in the $y - z$ plane. The θ variable is not invariant under boosts along the z axis, and so instead of the θ angle the *pseudorapidity*[2] η is used:

$$\eta = -\ln\left[\tan\left(\frac{\theta}{2}\right)\right] \tag{2.3}$$

[2]Actually the real boost-invariant variable is the *rapidity* y: $y = \frac{1}{2}\ln\frac{E+p\cos\theta}{E-p\cos\theta}$. In the ultra-relativistic limit the rapidity y can be substituted with the pseudorapidity η

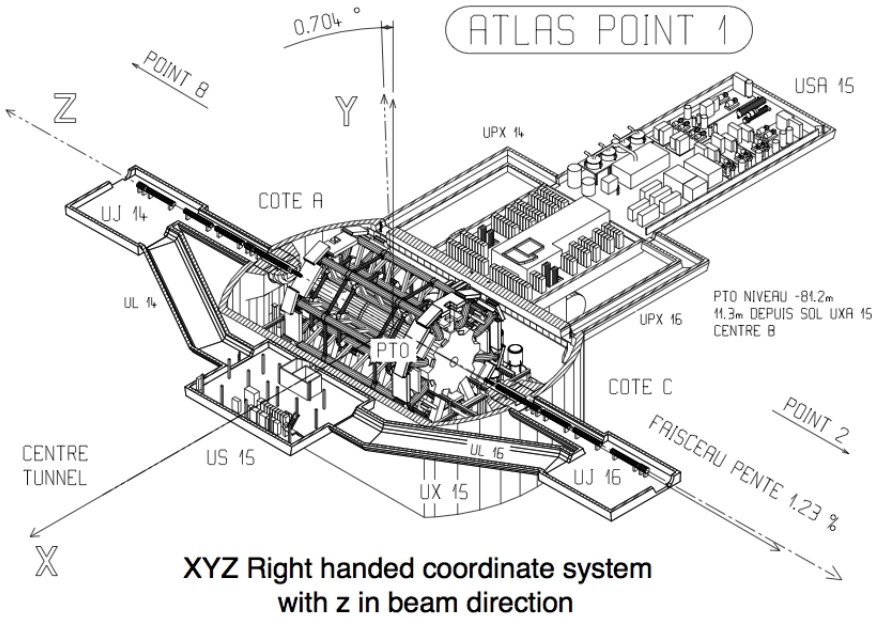


Figure 2.5 Reference system used in ATLAS.

Since at an hadronic collider the real colliding particles are the partons inside the protons, we can say that the actual center of mass energy in unknown in each collision: $\hat{s} = x_1 \cdot x_2 \cdot s$, where $\hat{s}$ is the effective collision energy, x_1 and x_2 are the fractions of momentum carried by the two colliding partons and s is the colliding energy of the two protons.

Because of this, the total momentum along the beam axis before the collision is unknown, while the total momentum in the transverse plane (i.e. the $x - y$ plane) is known to be zero and hence we can apply the momentum and energy conservation laws only on the transverse plane (because we know what is the initial total momentum).

For this reason transverse quantities are considered, and they will be denoted with the "T" sub-script (e.g. p_{T} stands for transverse momentum, that is the projection of the momentum on the $x - y$ plane).

2.2.2 ATLAS Magnets

The ATLAS detector is equipped with two magnetic systems: a superconducting solenoid [83], providing a magnetic field to the inner tracking system, and a system of air-core superconducting toroidal magnets [84, 85] located in the outer part of the detector as shown in Figure 2.6.

The solenoid covers the central region region of the detector, provides an uniform magnetic field of approximately 2 T along the z axis bending tracks of the particles

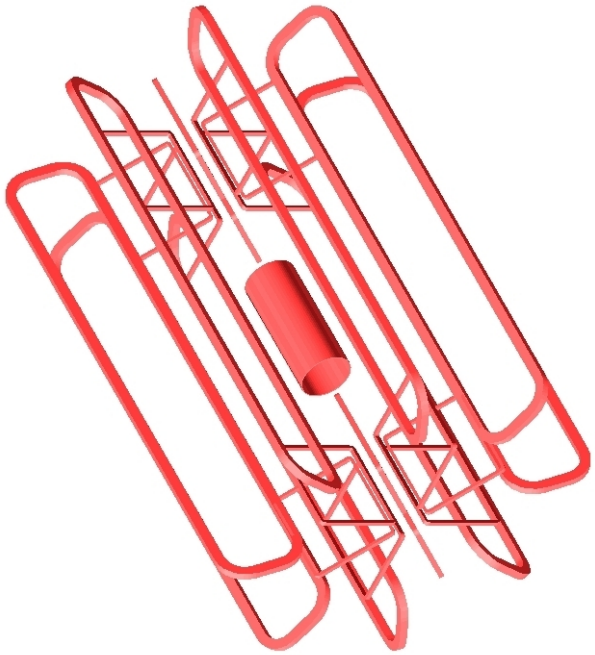


Figure 2.6 The magnetic system of the ATLAS detector: the inner cylinder is the superconducting solenoid, while the external parts are the coils of the toroid.

in the transverse plane in order to let the inner tracking system measure their transverse momentum.

The solenoid is located between the inner detector and the electromagnetic calorimeter and its dimensions (its width, particularly) have been optimized in order to minimize the amount of dead material (only 0.83 radiation lengths) in front of the calorimetric system.

The toroid is one of the peculiarities of the ATLAS detector: it is located outside of the calorimetric system covers the region $|\eta| < 3$ (considering all its subparts), and provides a magnetic field whose peak intensities are 3.9 T in the central region of the detector and 4.1 T in the forward region.

The aim of such a toroid is to have a large lever arm to improve the measurement of the muon transverse momentum, and it is built "in air" in order to minimize the muon multiple scattering within the detector.

The ATLAS double magnetic system has been designed to provide two independent measurements of the muon transverse momentum in the inner detector and in the muon spectrometer, thus ensuring good muon momentum resolution from few GeV up to the TeV scale.

2.2.3 The Inner Detector

The ATLAS Inner Detector tracker (ID), shown in Figure 2.7, is composed by three concentric cylindrical subdetectors. Its axis is centred on the z axis and it is approximately 6 m long and its diameter 2.30 m, covering the region ($|\eta| < 2.5$).

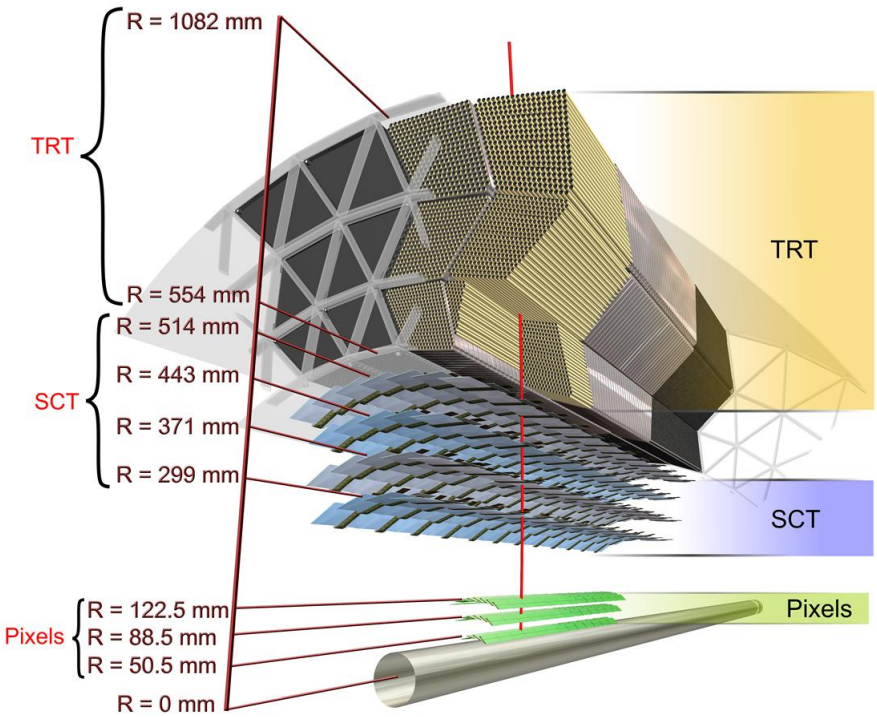


Figure 2.7 The ATLAS Inner Detector tracker: the three subdetectors (the Pixel Detector, the Semiconductor Tracker and the Transition Radiation Tracker) are shown as well as their radial dimensions.

The three sub-detector into the ID are:

- **Pixel Detector:** it is composed of three layers of silicon pixels. It provides high-precision track measurement, since the spatial resolution on the single hit is ~ 10 μm in the ϕ coordinate and ~ 115 μm along the z coordinate.
- **Semiconductor Tracker (SCT):** it is the second high-precision detector of the ATLAS inner tracker. It is composed of eight layers of silicon strips with a spatial resolution on the single hit of 17 μm in ϕ and 580 μm along z. The Pixel Detector and the Semiconductor Tracker together provide on average eight high-precision hits per track.
- **Trasition Radiation Tracker (TRT):** it is composed of straw tubes cham-

bers. The resolution of such a detector is lower than the previous one ($\sim 130\ \mu$m per straw), but it is compensated by the high number of points per track (36 on average) that it can provide.

The aim of the ATLAS ID is to measure the tracks of the charged particles produced in the pp collision and all the related features: p_{T}, η, ϕ, the eventual secondary vertexes due to long-lived particles.

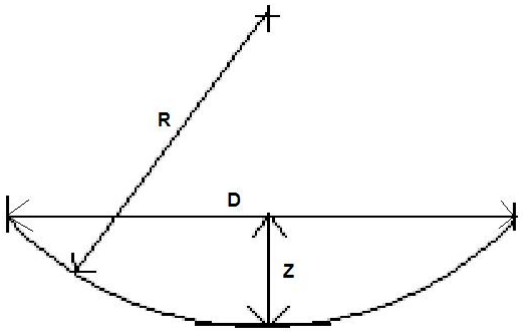


Figure 2.8 The sagitta of a track is the maximum distance between the track itself (that is an arc of a circle) and the straight segment having the same starting and ending points.

The momentum is measured by measuring the track curvature in the magnetic field provided by the superconducting solenoid described in the previous section. To estimate the expected resolution the *sagitta method* can be used: the magnetic field bends the trajectory of the charged particles in the ϕ coordinate because of Lorentz's force:

$$\vec{F}_L = q\vec{v} \times \vec{B} \tag{2.4}$$

where q is the charge of the particle, $\vec{v}$ is its velocity and $\vec{B}$ is the magnetic field. The resolution of the momentum measurement depends on many detector-related parameters:

$$\frac{\Delta p}{p^2} = \frac{8}{0.3 \cdot B \cdot L^2}\Delta s \tag{2.5}$$

where B is the magnetic field expressed in Tesla, L is the lenght of the reconstructed track expressed in meters, while Δs is (see Figure 2.8):

$$\Delta s = \frac{\epsilon}{8}\sqrt{\frac{720}{N+4}} \tag{2.6}$$

where N is the number of measured points on the track and ϵ is the resolution

on the measurement of the points.

From formulas 2.6 and 2.5 it is possible to see how it is crucial to have a strong magnetic filed, an high number of points per track and a good spatial resolution on these points in order to have a good resolution on the track p_{T}.

2.2.4 The Calorimetric System

In an high-energy physics experiment the calorimeters are used to measure the energy of photons, electrons (the electromagnetic calorimeter), hadronic jets (hadronic calorimeter) and the missing E_T (due to undetected particles like neutrinos) which is measured thanks to the tightness of the calorimetric system.

The ATLAS calorimeter has a cylindrical shape centred around the interaction point with its axis lying on the ATLAS z axis. It is long about $13m$ and the external radii of the electromagnetic and hadronic calorimeters are 2.25 and 4.25 m respectively.

The ATLAS calorimeters are represented in Figure 2.9 and the absorption lengths as a function of η are shown in Figure 2.10.

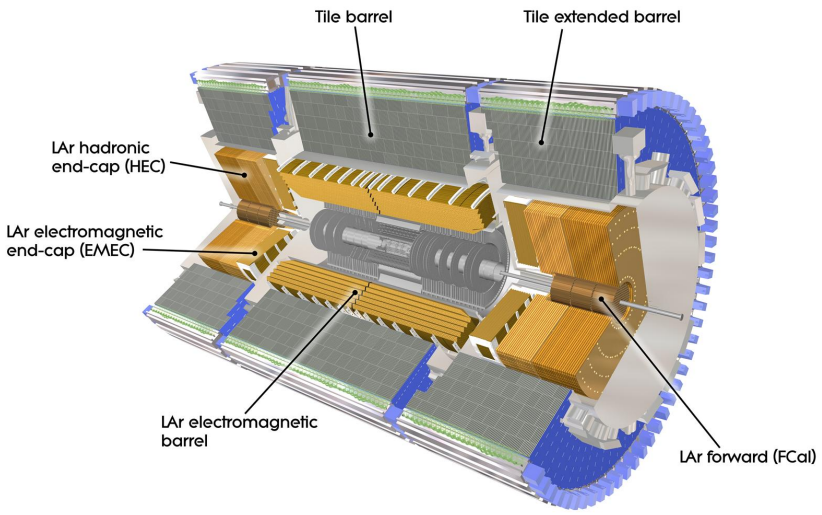


Figure 2.9 The ATLAS calorimetric system: the electromagnetic calorimeter made of liquid Argon and Lead and the hadronic caloimeter, whose composition varies as a function of η.

The Electromagnetic Calorimeter

The Electromagnetic Calorimeter of the ATLAS experiment covers the region up to $|\eta| < 3.2$. The structure of the Electromagnetic Calorimeter has a special feature, how you can see in Figure 2.11: it has an accordion structure made of lead (whose

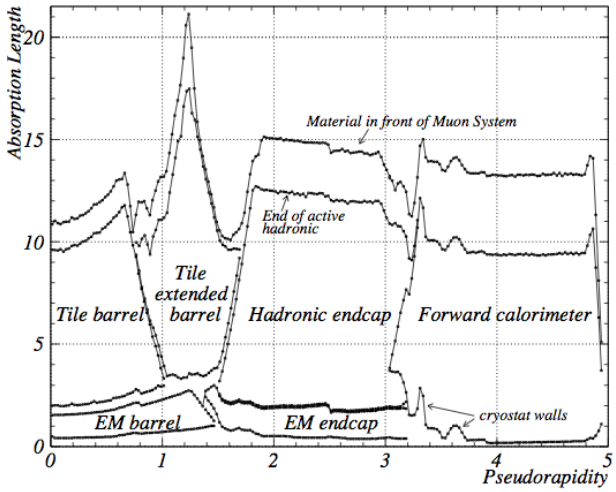


Figure 2.10 Amount of material in terms of absorption length in the ATLAS calorimetric system as a function of η.

thickness varies as a function of η in order to maximise the energy resolution) which is immersed in liquid Argon, which is the active material of the calorimeter. This structure confers to the calorimeter very high acceptance and symmetry in the ϕ coordinate.

In the central region $|\eta| < 2.5$ the radial coordinate the electromagnetic calorimeter has three sampling channels in order to maximize particle identification power (see Figure 2.11).

The calorimeter is segmented in cells of variable dimensions as a function of η as well as its thickness ($> 24X_0$ in the central region and $> 26X_0$ in the forward region): in the central region the segmentation is $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$.

The ATLAS EM calorimeter energy resolution is parametrized as:

$$\frac{\Delta E}{E} = \frac{10\%}{\sqrt{E[\text{ GeV}]}} \oplus 1\% \tag{2.7}$$

Where 10% is the sampling term and 1% is the constant inter-calibration term. The η resolution is:

$$\frac{40\text{mrad}}{\sqrt{E[\text{ GeV}]}} \tag{2.8}$$

The Hadronic Calorimeter

The Hadronic Calorimeter covers the region $|\eta| < 4.5$, and it is realized with a variety of techniques as a function of η like it is possible to check in Figure 2.9.

The central region ($|\eta| < 1.7$) it is made of alternating layers of iron (used as absorber) and scintillating tiles as active material, and its thickness offers about 10

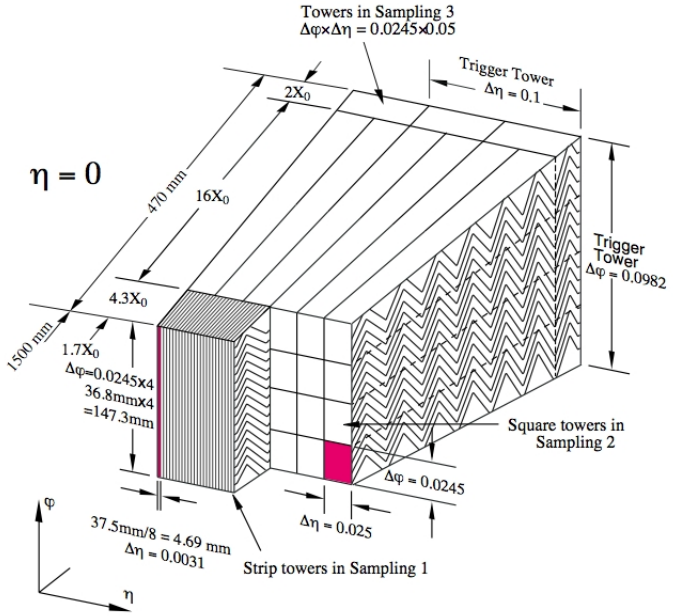


Figure 2.11 The accordion structure of the electromagnetic calorimeter and its radial segmentation.

interactions lengths λ at $\eta = 0$. It is segmented in $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ pseudo-projective towers pointing to the interaction point.

The "endcap" region ($1.7 < |\eta| < 3.1$) is equipped with a liquid Argon and lead, as the Electromagnetic Calorimeter, while the forward region ($3.1 < |\eta| < 4.5$) is equipped with liquid Argon, but the accordion structure is replaced by a concentric rods and tubes made of copper. This variety of materials and structures is due to the different radiation hardness required in the different parts of the detector.

2.2.5 The Muon Spectrometer

The ATLAS Muon Spectrometer (MS) is instrumented with both trigger and high-precision chambers immersed in the magnetic field provided by the toroidal magnets which bends the particles along the η coordinate, and it allows to measure the muons p_{T} in the region $|\eta| < 2.7$ using the sagitta method described in section 2.2.3. Here the lenght of the lever arm plays a leading role on the p_{T} resolution. The MS is shown in Figure 2.12.

The chambers used to reconstruct the muon track are of several types depending on the η region, in order to face the different rate conditions present in the different parts of the detector. In the central region ($|\eta| < 2$) Monitored Drift Tubes (MDTs) are used. The MTD chambers are composed of aluminium tubes of 30 mm diameter and 400 μm thickness, with a 50 μm diameter central wire. The tubes are filled with a mixture of Argon and CO_2 at high pressure (3 bars), and each tube has a spatial resolution of 80 μm.

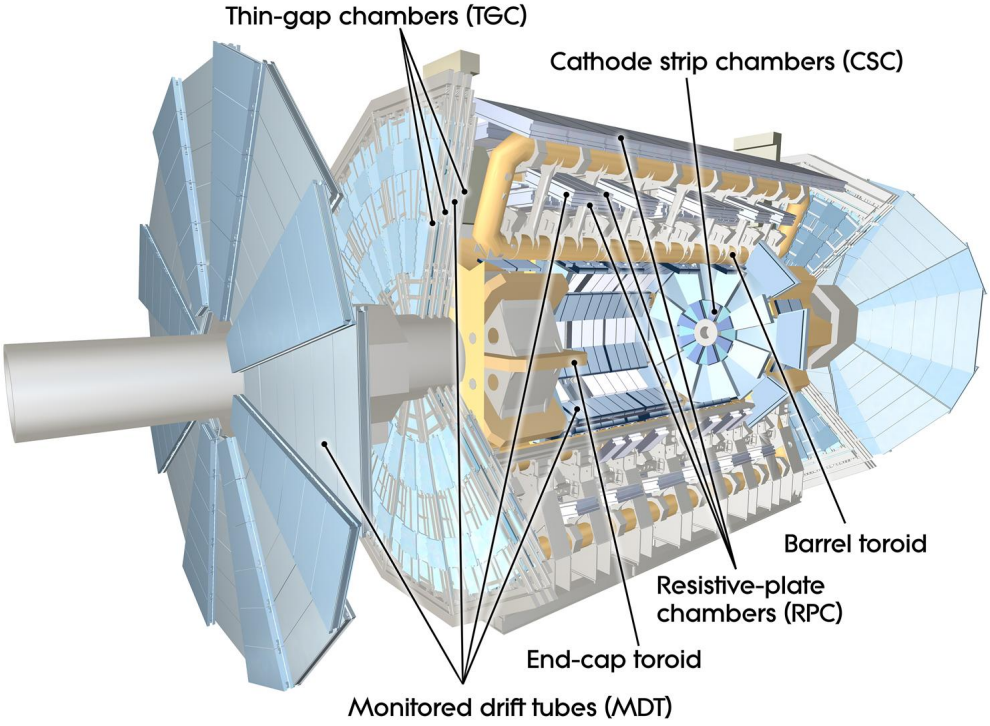


Figure 2.12 The ATLAS Muon Spectrometer.

At higher pseudo-rapidity ($2 < |\eta| < 2.7$) the higher granularity of the Cathode Strip Chambers (CSC) are used. CSC chambers are multiwire proportional chambers in which the readout is performed using strips forming a grid on the cathode plane in both orthogonal and parallel direction with respect to the wire. The spatial resolution of the CSC is about 60 μm.

As shown in Figure 2.12, in the central region the MS is arranged on a three layer −or stations− cylindrical structure which radii are 5, 7.5 and 10 m; while in the forward region the detectors are arranged vertically, forming four disks at 7, 10, 14 and $21 - 23$ m from the interaction point.

The other chambers installed on the spectrometer are used for the trigger (see next section for details). The chambers used for the muon trigger are Resistive Plate Chambers (RPC) in the central region ($|\eta| < 1.05$) and Thin Gas Chambers (TGC) in the forward region.

These detectors provide very high time resolution ($\mathcal{O}$(ns)) even if the spatial resolution is not so high ($\mathcal{O}$(cm)).

2.2.6 The ATLAS Trigger

The LHC is designed to provide collisions at a frequency of 40 MHz and, since the average dimension of an ATLAS event is ~ 1.5 MB, a recording rate of ~ 60 TB per second would be needed, while the current technology allows to record data at about 300 MB/s. To deal with this environment and knowing that the interesting physics at LHC occurs at very low rate, as shown in Figure 2.14, the events to be recorded can be selected without loosing the relevant information.

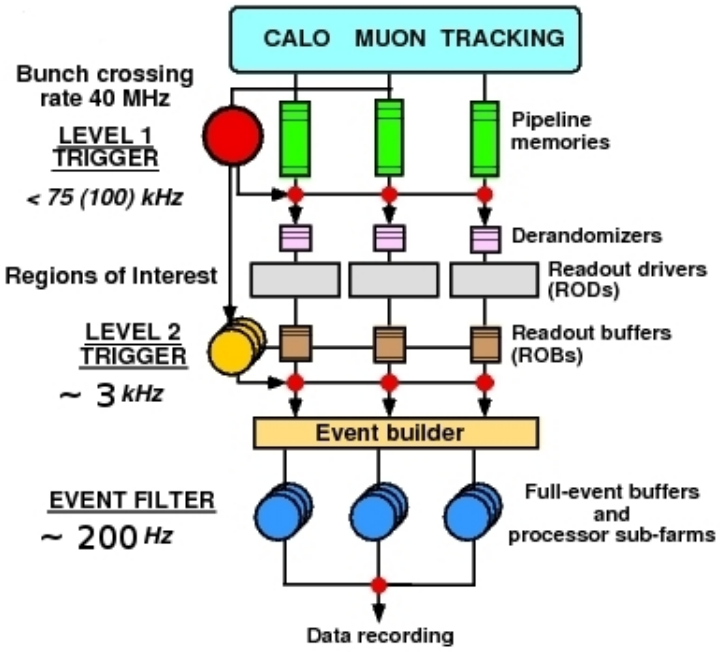


Figure 2.13 Main structure of the ATLAS trigger system: it is made of three levels, each improving the measurement of the previous levels also combining informations from different subdetectors.

This selection is performed on-line by the ATLAS trigger and data acquisition system [86]. The ATLAS trigger is designed to rapidly inspect the events detected by the ATLAS detector and choose whether record or discard the event after having compared its main features with a set of predefined thresholds contained in the trigger menu. In case that the trigger decides to discard an event, then the event is not recorded and lost forever.

The ATLAS trigger system has a three level structure: each level refines the measurements of the previous level introducing also new selection criteria and combining the information from different subdetectors, as shown in Figure 2.13.

The first level of the ATLAS trigger (L1 or LVL1) is completely hardware-based and it makes use of only the data collected by the calorimetric system and the muon spectrometer: the L1 trigger only looks for high-p_{T} muons candidates or calorimetric objects (electrons/γ, jets) by means of fast and rough measurements performed

by ad-hoc detectors in the Muon Spectrometer (RPC, TGC) and simplified object identification in the calorimeter.

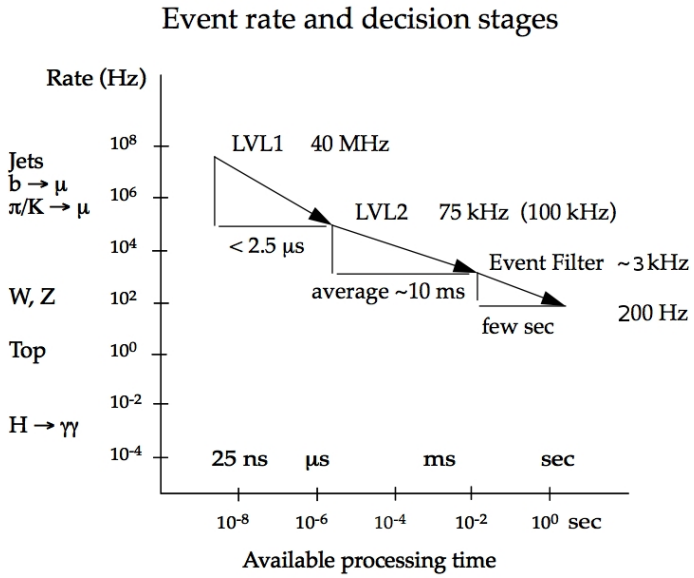


Figure 2.14 The event rate at which interesting physics occur −referred to LHC design parameters− and the processing time of each trigger level.

The L1 is designed to take a decision on the event in 2.5 μs and its output is a list of so-called *Regions of Interest* (RoI), which are $\eta - \phi$ regions of the detector in which interesting activity has been detected, and the output rate is about 100 kHz.

The second level of the ATLAS trigger (L2 or LVL2) is completely software-based. It takes as input the RoIs provided by the L1, and refines the measurement into these regions: data of the precision chambers are used in the Muon Spectrometer (MDT, CSC) as well as the data from the ID, while the measurement of the calorimetric objects is refined using higher level algorithms.

Moreover the data of the different subdetectors are combined together in order to obtained better object reconstruction/identification (e.g. the ID and the MS tracks are combined for the muons, ID and calorimetric informations are combined to discriminate between electrons and photons). The L2 takes its decision in $\mathcal{O}$(10ms) and its output rate is about 3 kHz.

The third level of the ATLAS trigger (Event Filter, EF) is completely software-based and forms, together with the L2, the High Level Trigger (HLT). At this stage a full reconstruction of the detector is performed (the measurement is not restricted to the RoIs), and the algorithms run at the EF are mostly the off-line reconstruction algorithms adapted to the on-line environment. The decision of the EF is taken in $\mathcal{O}$(1s) and the output rate is about 400 Hz.

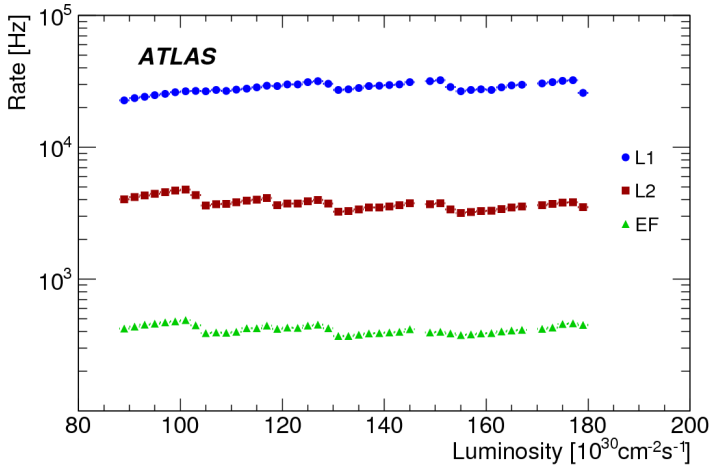


Figure 2.15 Total trigger rates at each level of the ATLAS trigger.

Figure 2.15 shows the total trigger rate for all the three levels as a function of the instantaneous luminosity: how can be seen the trigger rates are kept stable. This happens thanks to changes in the prescales and in the trigger menu −on-line into the ATLAS Control Room−, where higher thresholds or quality criteria on the trigger objects are required as the luminosity increases.

Electron Trigger

The electron trigger follows the three level ATLAS trigger structure, in which the measurements and the selections are refined at each stage. At the first level the electron trigger makes use only of the calorimeters, and hence no distinction between electrons and photons is possible since they are both identified as "calorimetric objects". In particular the L1 trigger measurement is a real calorimetric measurement even if it is done with reduced granularity, represented in Figure 2.16:

Once a relevant amount of energy is detected, the total energy in a little 2×2 cluster is measured (green area), and the isolation with respect to electromagnetic (yellow area) and hadronic activity (pink area, e.g. due to electrons coming from heavy quark decay) is computed. If the these three parameters (E_T, electromagnetic and hadronic isolation) fulfil the requirements, then the electromagnetic calorimeter is accepted as a good calorimetric object and its RoI is propagated to the L2.

The L2 trigger basically refines the calorimetric measurement, accessing the full granularity of the calorimeters and studying the shape of the energy deposit (e.g. π^0/γ separation), and includes the data of the inner tracking system. At this level a "calorimetric object" may become an electron if an ID track consistent with it is found. Since the measurements are more precise at this level, tighter conditions on the quality and the kinematic features of the electron candidates can be required.

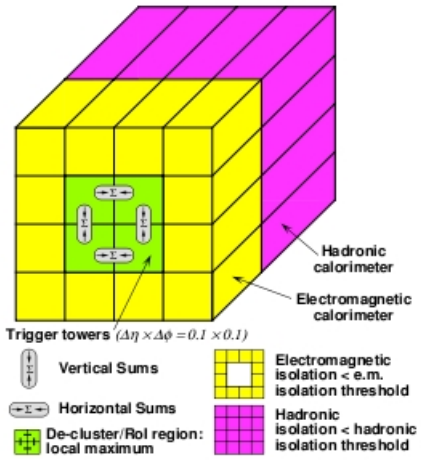


Figure 2.16 The L1 trigger for calorimetric objects in the Electromagnetic Calorimeter: the green area represents the RoI cluster, the yellow area is the region used for the isolation requirements, and the pink area is the region used for the hadronic isolation.

At the end of the chain the EF further refines the measurements performed at the L2 on the electron candidates, running algorithms very similar to the off-line ones and having access to the data of all the subdetectors with full granularity.

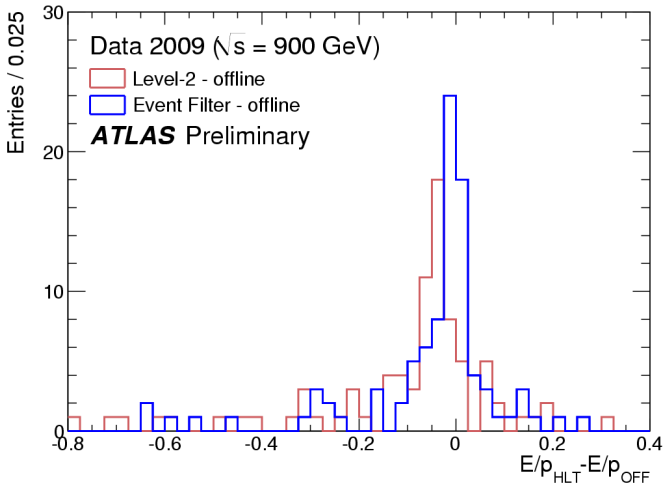


Figure 2.17 E/p distribution found by the HLT and the off-line for the electron trigger. The distributions are shown for L2 and EF separately.

The distribution of the difference between the off-line and the value measured at different trigger levels of the E/p variables for electrons is showed in Figure 2.17. This shows how the EF measurement (blue line) is better than the L2 measurement (red line), since the former is allowed to use reconstruction algorithms very similar to the off-line ones thanks to the large processing time available (see Figure 2.14),

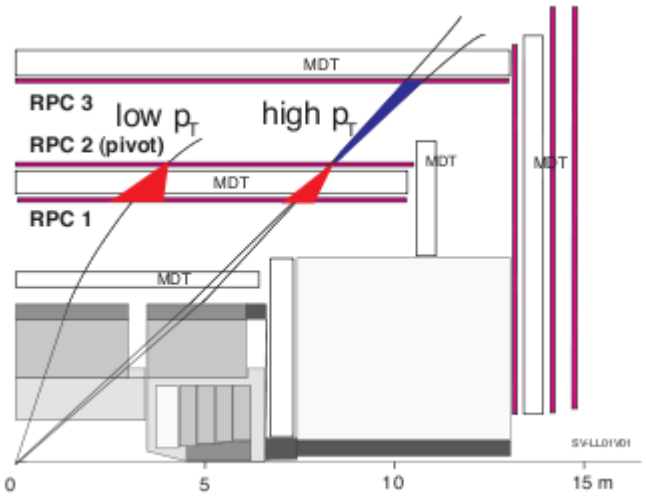


Figure 2.18 L1 muon trigger algorithm: a muon coming from the interaction point leaves hits on the three layers of RPC detectors installed in the muon spectrometer. The position of the different hits is correlated as a function of the muon p_{T}.

while the latter has to rely on simplified algorithms.

Muon Trigger

The L1 muon trigger relies on the temporal and geometric correlation of the hits left by a muon on the different layers of RPC detectors installed in the muon spectrometer, as shown in Figure 2.18.

When a muon coming from the interaction point crosses the RPC detectors, it leaves hits on each of them: starting from the hit on the central station (also known as *pivot* plane, RPC2 in Figure 2.18) a "correlation window" (several windows are opened for several p_{T} thresholds) is opened on the RPC1 layer.

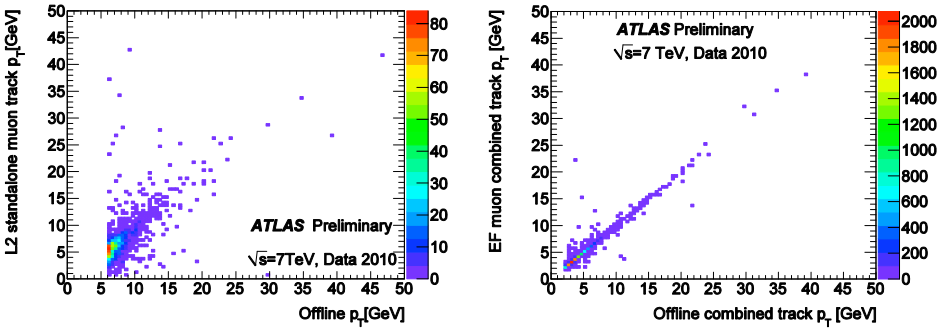
If a good hit (i.e. hits in both η and ϕ and in time with the hit on the pivot plane) is found on the RPC1 layer then a low-p_{T} muon candidate is found. The same algorithm is applied using the RPC3 plane to look for high-p_{T} muon candidates. Once a muon candidate is found, the RoI is propagated to the L2.

At the L2 the muon track is reconstructed for the first time: there are algorithms which reconstruct the muon tracks in the ID and in the MS separately and then combine them in order to determine of p_{T}, η and ϕ. At this level the p_{T} measurement is not done by a fit, but *look-up tables* are used: the p_{T} estimation is done starting from the relation

$$\frac{1}{s} = A_0 \cdot p_{\mathrm{T}} + A_1 \tag{2.9}$$

where s is the sagitta of the muon track and A_0 and A_1 are two constant values needed to take into account the magnetic field and the energy loss in the calorimeters respectively. A look-up table is basically a table whose columns and rows represent the $\eta - \phi$ segmentation of the ATLAS detector, and in each cell a (A_0, A_1) pair is contained. For each muon candidate, given η, ϕ and s, a fast estimation of the p_{T} is possible. This method is used since at the L2 there is not enough time to perform a real fit to precisely measure the track p_{T}. Once the full track is reconstructed (from the ID to the MS), the calorimetric activity around it is measured, in order to apply the isolation requirements.

At the EF the muon reconstruction algorithms perform again the operations performed by the L2 algorithms, but now the full detector with its full granularity can be accessed, and a real fit of the muon track is performed.



(a) (b)

Figure 2.19 Correlation between the muon p_T reconstructed at several trigger levels (level 2 in (a) and event filter in (b)) and the off-line reconstruction.

Figure 5.15(a) shows the correlation between the muon p_T reconstructed at different trigger levels and the off-line reconstruction: in Figure 2.19(a) the correlation between the L2 stand alone p_T is shown, while in Figure 2.19(b) the correlation between the EF combined p_T measurement and the off-line one is shown.

As can be seen the EF measurement is much more accurate and precise compared to the one performed at L2. The corresponding plot for L1 is not shown since at L1 the muon p_T is not really measured, but, as explained above, only a threshold is available.

2.2.7 Monte Carlo Generators in ATLAS Event Simulation

Adding to the real events recorded by the ATLAS detector, the use of the Monte Carlo simulations is essential for the different physics programs.

During the preparation phase of an experiment, simulation provides the environment to develop and understand the detector, to develop analysis strategies, to estimate the sensitivity to different physics processes, to develop and validate

object reconstruction algorithms, to optimize the trigger menus, and so on. During the running phase of an experiment, simulation is used to compare predictions of theoretical models against the real data. The event simulation and reconstruction is performed in the Athena frame-work[39], who is in charge of the following processes:

- *Event generation*: corresponds to the phase of proton-proton (pp) collision events generation. It takes care of the production and decay of particles in a given process. Several event generators are available.
- *Detector simulation*: is the simulation of interactions between the generated particles and the detector.
- *Digitization*: corresponds to the simulation of the detector readout, or better says, the conversion of energy deposited in the detector to times, currents and voltages for readout electronics. The output format of the simulation is identical to the real detector output format.
- *Reconstruction*: in this step a set of object reconstruction algorithms are applied. These algorithms are applied to both simulation and real data in exactly the same way. This phase is better described in Chapter 3.

There are two kinds of Monte Carlo (MC) generators, the so-called *multi-propuse MC* generators which handle all the event generation steps, and the *specialized MC* generators that handle only some specific steps of the MC generation chain.

The kinematic distributions for a given process may differ between different Monte Carlo generators. Depending on the problem under study, one generator may be more suitable than others and the comparison between different generators is always encouraged.

The Monte Carlo generators used into this Higgs search analysis are listed below, together with their main characteristics.

- Pythia: is a multi-purpose Monte Carlo generator for event simulation in pp, e^+e^- and ep colliders. Pythia simulates non-diffractive proton-proton collisions using a $2 \to n$ ($n \leq 3$) matrix element at LO to model the hard subprocess, and uses p_T-ordered parton showers to model additional radiation in the leading-logarithmic approximation. The hadronisation model used is the Lund string model. MPIs are also simulated[35].
- Herwig: is a general purpose Monte Carlo generator, which uses a LO $2 \to 2$ matrix element supplemented with angular-ordered parton showers in the leading-logarithm approximation. The cluster model is used for the hadronisation. The underline is modelled using an external package called Jimmy [36, 37, 38].
- Herwig++: is based on the event generator Herwig, but redesigned in the C++ programming language (Herwig is programmed in Fortran). The generator

contains a few modelling improvements. It also uses angular-ordered parton showers, but with an updated evolution variable and a better phase space treatment. The cluster model is also used for hadronisation. The UE are described using a multiple partonic interactions model[40, 41].

- Alpgen: is a specialized tree matrix-element generator for hard multi-parton processes $2 \to n$ ($n \leq 9$) in hadronic collisions. It is interfaced to Herwig to produce angular-ordered parton showers in leading-logarithmic approximation or Pythia to produce p_T-ordered parton showers. Parton showers are matched to the matrix element with the MLM matching scheme. The hadronisation process is simulated with Herwig, using the cluster model. MPIs are modelled using Jimmy[42].

- MC@NLO: is a Fortran package which allows to match NLO QCD matrix elements consistently into a parton shower framework. In order to reproduce the NLO corrections fully, some of the configurations have negative weights. The shower and hadronization can be implemented using Herwig or Herwig++. The NLO expansion of the hard emissions needs to be evaluated for each showering program used[43, 44].

- Powheg: is a parton-level Monte Carlo generator. It allows to interface NLO calculations with a parton shower framework. It generates the hardest emission first, with NLO accuracy independently of the parton shower generator used. It can be interfaced with several parton shower generators as Herwig, Pythia, etc[45].

Chapter 3

Physiscs Objects Reconstruction in ATLAS

The outputs of the digitization process of the detector signals generated by the particles produced in real/simulated pp collisions are processed by a serie of algorithms in order to built physics objects up. This step is not done in real time and thus is known as off-line event reconstruction. The result is a set of physics objects with four-momenta that can be used directly in physics analyses. The reconstruction of tracks, primary vertices, electrons, muons, neutrinos and jets as well as the trigger chains used in the different analyses of this thesis will be described in the following sections.

3.1 Data quality

The events where some of the relevant ATLAS subdetectors were not properly operational can not be used for physics analyses. In ATLAS, each subsystem is in charge of setting its own data quality and integrity flags for each Luminosity Block (LB) [1]. This information can be used to create a list of LB usable for analyses, called Good Runs List (GRL). Each analysis uses a GRL to reject those events affected by issues in the relevant subdetectors.

A special case during the 2011 created an exception in the data quality assessment was the LAr hole issue. The information of six front end boards in the LAr calorimeter was lost due to a problem with their controller board. It created a "hole" in the detector data collection that affected approximately 948.6 pb^{-1} of data. In release 16 of the ATLAS software, this issue was not modelled in the simulation. Instead of removing all the events affected, as would have been the procedure using a GRL, it was decided to remove only those events where the object reconstruction was affected by the issue. In release 17 of the ATLAS software, the issue was simulated in the Monte Carlo samples used. Therefore, no correction for the acceptance loss was needed.

[1] A luminosity block is the unit of time for data-taking, and lasts about two minutes.

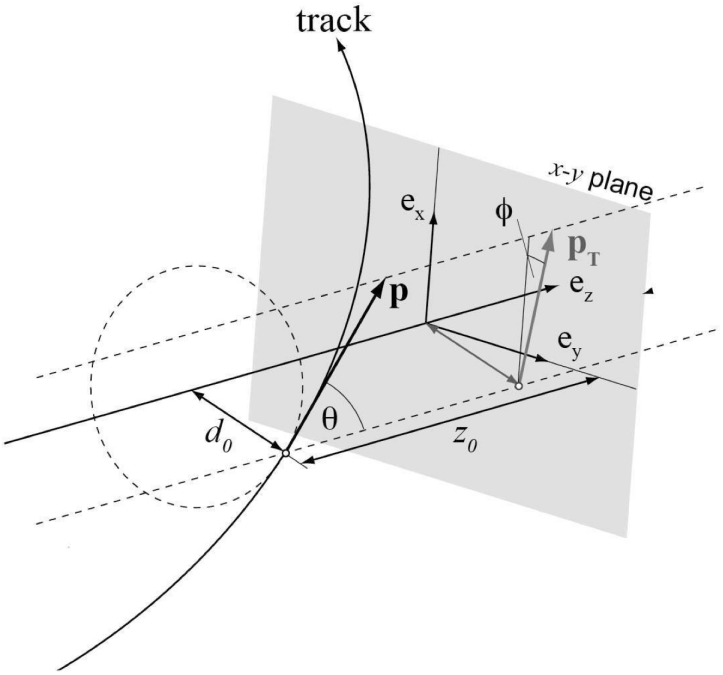


Figure 3.1 A track can be parametrized at one given point by its position, transverse momentum and charge. The position at any other point can be calculated if the magnetic field and the detector material is known.

3.2 ID Tracks

Tracks represent the trajectory of charged particles inside the detector. They are reconstructed using information from the ID, documented in Section 2.2.3. A precise track reconstruction is important to achieve a high vertex reconstruction efficiency and high precision in the particle momentum measurement.

In ATLAS, tracks are parametrized by five parameters defined at the track's trajectory point closest (unless otherwise stated) to the center of the beam-spot: d_0, and z_0 (radial and longitudinal impact parameters), ϕ and θ (azimuthal and polar angle) and $\frac{sign(q)}{p_t}$ (Figure 3.1). The beam spot is the region where both beams interact, which does not correspond exactly to the geometrical center of the ATLAS detector.

Lets summarized the three main steps of the track reconstruction[46]:

- **Track finding:** assignment of ID hits to track candidates.
- **Track fit:** determination of track parameters and their errors. Provides track's fit quality variables.
- **Test of track hypothesis:** check the track candidate quality and the overlap with others tracks candidates.

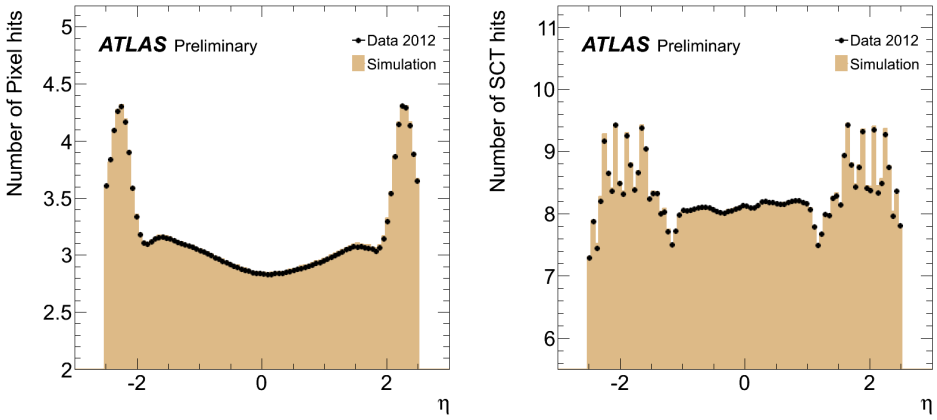
In ATLAS the track finding and fit steps are merged. The hits from the pixel detector and the SCT are transformed into three-dimensional space points. Just after, the inside-out algorithm[47] take action. It consists in adding the three-dimensional space points one by one moving away from the interaction point to form a "road". It uses a combinatorial Kalman filter[48]. The Kalman filter algorithm adds three-dimensional space points iteratively and fits simultaneously the track candidate. Cuts on the quality of the fit are applied to eliminate poor quality tracks and to avoid overlaps with others tracks candidates. The selected tracks are extended into the TRT and fitted again to get the final values of the track parameters.

The mean energy loss in the detector material, the multiple scattering, the Bremsstrahlung effect and the changes in the magnetic field along the track trajectory are taken into account during the track fitting process. The inside-out algorithm provides the best reconstruction efficiency of primary charged particles directly produced in a pp collision or from decays or interaction of particles with a short lifetime ($< 3 \times 10^{-11}$s).

In order to better reconstruct secondary charged particles, produced in the interaction of primaries (with a lifetime $> 3 \times 10^{-11}$s), or conversion candidates an additional track finding algorithm, called outside-in, is applied. The track finding process starts with TRT segments not used by the inside-out algorithm. They are then extended to the SCT and pixel detector.

During 2011 data taking the detector occupancy increased significantly. Under these conditions the possibility of having incorrect hits assignments and more fake tracks from random hit combinations increase. The performance of the track reconstruction at ATLAS has been recently studied in the 2011 high pile-up environment[48].

The efficiency remains almost unchanged. However, the fraction of combinatorial fake tracks increases with the average number of pp collisions per bunch crossing at the time of the recorded event $< \mu >$.



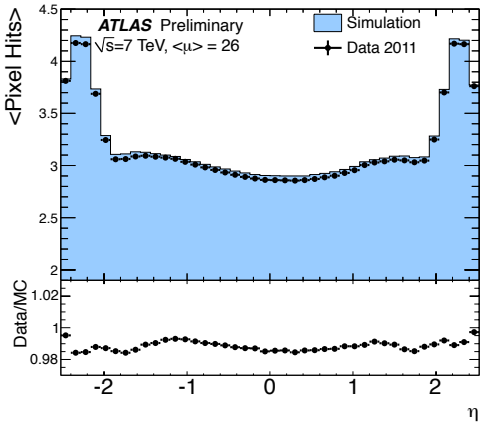


Figure 3.2 (a) 2012 data and simulation comparison of Pixel hits vs eta, (b) 2012 data and simulation comparison of SCT hits vs eta. (c) The number of pixel, SCT and TRT hits as a function of η for tracks with 2011 data.

3.3 Primary vertices

Two different kinds of vertices can be present in an event:

- **The primary vertices (PV)** which correspond to the collision point of beam particles. They are characterized by having many associated tracks-particle, thus a high track multiplicity. In a selected event usually there is one hard-scatter PV, while the rest are associated, by conventions, to pile-up interactions.

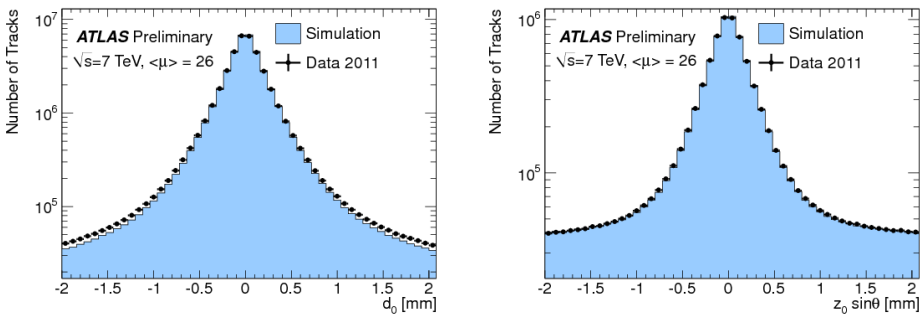
- **The secondary vertices** correspond to decay of short-lived particles, which decay at a measurable distance from the PVs. The track multiplicity for secondary vertices is lower.

In general, the PV reconstruction is done in three steps:

- **Vertex finding:** assignment of reconstructed tracks to PV candidates.
- **Vertex fit:** reconstruction of the PV position, calculation of its error matrix, estimation of the fit quality and optional refit of the associated track's parameters to constrain them to originate from the corresponding PV and not from the beam spot.
- **Test of vertex hypothesis:** check the vertex candidate quality and the overlaps with others vertices candidates.

Into the ATLAS experiment, the PV reconstruction is done using an iterative vertex finding algorithm[49]. First, a vertex seed is found by looking for the maximum of the z_0 tracks distribution. An iterative χ^2 fit is used to fit the seed and the surrounding tracks. The matrix errors of the tracks are properly taken into account during the vertex fit. Tracks incompatible with the PV candidate (displaced by more than 7σ from the vertex) are used to seed a new PV. This procedure is repeated until no unassociated tracks are left or no additional vertex can be found. PVs are required to have at least two associated tracks. The same track can be associated to multiple vertices. The PV with the largest sum of squared transverse momenta $\sum p_T^2$ of the tracks is chosen as the hard-scatter PV. Figures 3.2 and 3.3 show the performance of ATLAS relative to reconstructions tracks, pixel and vertex detector-variables.

In high pile-up environment, the increasing number of fake tracks increases the probability to reconstruct a fake vertex. Furthermore, the common presence of nearby interactions increases the probability of reconstructing only one vertex out of several. Studies using 2011 data have shown that the PV efficiency reconstruction decreases with increasing $< \mu >$[48]. Some quality criteria are applied to the tracks used in the PV reconstruction process. They vary from one analysis to the other.



(a) (b)

Figure 3.3 (a) The impact parameter distributions with respect to the primary vertex of tracks and the corresponding simulation sample, (b) the impact parameter distributions with respect to the primary vertex of tracks and the corresponding simulation sample. The simulation in each of the plots has been reweighted to match the pT spectrum of the data.

3.4 Electrons

3.4.1 Electron reconstruction

The distinguishing signature of an electron is a curved track in the inner detector and a narrow shower in the EM calorimeter. Electrons can be produced in the hard-scatter interaction, but also inside jets or in photon conversions. Then, there is a probability of misclassifying narrow jets or photons as electrons. During the electron reconstruction and identification, several criteria are applied to determine if a given energy deposit and associated track were produced in the hard-scatter interaction or not. The electrons in the central region ($|\eta| < 2.47$) are reconstructed using an algorithm that combines the information from the EM calorimeter and the ID. Other algorithms reconstruct electrons in the forward region using only information from the EM calorimeter. In this thesis only electrons in the central region will be used[50]. Therefore, the reconstruction of electrons in the forward region will not be discussed[51].

Electron reconstruction in ATLAS in the central region is done using the sliding-window algorithm[52]. It searches for clusters in the middle layer of the EM calorimeter with a total $E_T > 2.5$ GeV. The window used to defined the clusters has a size of 3×5 in middle layer cell units ($\Delta\eta \times \Delta\phi = 0.025 \times 0.025$). If the seeded cluster is matched to a pair of tracks originating from a reconstructed photon conversion vertex, it is tagged as a converted photon. Otherwise, if the algorithm is able to match a track from the ID with the seeded cluster, it is tagged as an electron candidate.

The matching is done in an $\Delta\eta \times \Delta\phi$ window of 0.05×0.10, to account for bremsstrahlung losses. The track momentum is required to be compatible with the cluster energy. In case that several tracks are matched to the EM cluster, the tracks with hits in the silicon detectors are preferred and the closest in ΔR is chosen. In addition, information from the TRT can be used to enhance the separation

of electron candidates from pions. The final clusters are built around the seeded clusters matched with a track, by including all cells from different EM calorimeter layers located inside a rectangle centred on the seed position. The rectangle size depends on the position in the calorimeter of the seeded clusters (barrel or endcap). The energy of the cluster is calibrated to the EM energy scale, which was derived from MC based corrections to account for energy loss in passive material, test-beam measurements, and measurements of $Z \rightarrow ee$ decays for final calibration[53, 54].

3.4.2 Electron identification

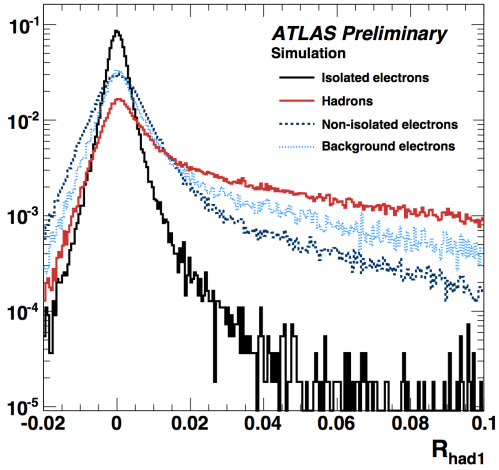
Electrons can be distinguished from hadrons since EM showers deposit most of their energy in the second layer of the EM calorimeter. The width of electron showers is narrower than for hadrons. The ratio of the transverse energy reconstructed in the first layer of the hadronic calorimeter to the transverse energy reconstructed in the EM calorimeter, known as **hadronic leakage** or R_{had}, is smaller for electrons than for hadrons. Also the ratio of the energy reconstructed in the EM calorimeter to the track momentum $\frac{E}{p}$ can be used as a discriminant variable, since it is smaller for charged hadrons (Figure 3.4).

The most difficult task is to distinguish electrons from π^0's and η's. They decay into two photons which form two close EM showers indistinguishable in the second EM calorimeter layer. In this case, the first layer of the EM calorimeter can be used due to its high granularity, to identify the two maximum in the π^0 or η shower corresponding to the two photons.

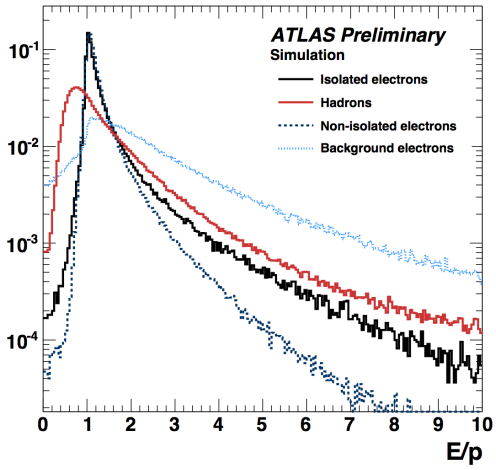
In ATLAS, there are six different series of cuts used in the electron identification process that provide good separation between electrons and jets faking electrons: *loose, loose++, medium, medium++, tight* and *tight++*[52].

In general, each cut adds to the previous some additional requirements. The "++" menu was incorporated starting from release 17, in order to accomplish the trigger bandwidth restrictions for high luminosity. The discriminating variables used are defined using calorimeter and ID information.

- **Loose:** is based on calorimeter information only. It requires electron candidates with $|\eta| < 2.47$ with low hadronic leakage and cuts on shower shape variables, derived from the energy deposits in the second layer of the EM calorimeter. The loose criteria provides a high identification efficiency (close to 95%). But the expected background rejection (the jet rejection), is low, about 500, i.e. one in 500 jets will pass the Loose selection.
- **Loose++:** adds additional cuts to the loose selection. It adds requirements related to the matched track: at least 1 hit in the pixel detector, at least 7 hits from both the pixel and SCT and the distance in $|\eta|$ between the cluster and the extrapolated track in the first EM layer has to be smaller than 0.015. Its efficiency, measured in $Z \rightarrow ee$ events, is close to the loose one ($93\% - 95\%$) with a higher expected rejection of about 5000[55].



(a)



(b)

Figure 3.4 (Example of some electron identification variables: (a) "hadronic leakage" R_{had}, (b) ratio of the energy reconstructed in the EM calorimeter to the track momentum $\frac{E}{p}$.

- **Medium:** adds additional criteria related to the shower shape calculated using the first EM layer and to the deviation in the energies of the largest and second largest deposits in this layer, allowing discrimination against π^0's and η's. In addition, the absolute value of the transverse impact parameter of the track, $|d_0|$, is required to be lower than 5mm and the distance in $|\eta|$ between the cluster and the extrapolated track in the first EM layer lower than 0.01. Its efficiency is of about 88%[55] and has a rejection higher than the one achieved by the loose++ selection.

- **Medium++:** requires at least one hit in the B-layer (first pixel detector layer) to reject electrons from photon conversions. Tracks having a low fraction of high-threshold TRT hits are rejected to decrease the contamination from

charged hadrons. $|\Delta\eta|$ between the cluster and extrapolated track in the first EM layer is lowered to 0.005. It has an efficiency of around 85%, with a expected rejection closer to 50000.

- **Tight:** requires that $|\Delta\phi|$ and $|\Delta\eta|$ between the cluster and the matched track has to be less than 0.02 and 0.005, respectively. A requirement on $\frac{E}{p}$ is introduced. The $|d_0|$ requirement is tightened (to be less than 1 mm), as well as the fraction of high-threshold TRT hits. The identification efficiency is around 75%, with a rejection higher than the medium++ one.

- **Tight++:** only adds asymmetric $\Delta\phi$ track-cluster matching cuts. It has an efficiency slightly better than the one for tight selection and a slightly better rejection too.

The performance of electron reconstruction, trigger and identification was evaluated with 2010 data and MC using $Z \to ee$ and $W \to e\nu$ events[52].

To suppress the background due to non-prompt leptons, e.g. from decays of hadrons (including heavy flavour) produced in jets, the leptons in the event are usually required to be isolated. A calorimeter isolation, a track isolation or both can be applied. The calorimeter isolation is estimated using the energy in a cone of $R = 0.2$ centred around the electron after the subtraction of the energy associated with the electron itself (`EtCone20`).

Track isolation is calculated using the scalar sum of tracks p_T in a cone of $\Delta R = 0.3$ centred around the electron without including the electron p_T itself (`PtCone30`). The calorimeter isolation variables usually include a correction for the increase in the energy of the electron in the isolation cone with electron p_T (transverse shower leakage) and for additional energy deposits from pile-up events.

3.4.3 Electron scale factors and energy corrections

In release 16 used in 2010, electron identification scale factors were calculated only as a function of η cluster. The differences in trigger and reconstruction efficiency were taken into account using flat scale factors of 0.995 ± 0.010 and 1.013 ± 0.015, respectively[52]. They have been estimated using tag-and-probe methods in $Z \to ee$ and $W \to e\nu$ samples.

In release 17 used in 2012, trigger, reconstruction and identification (including isolation) efficiency scale factors for electrons (with its respective systematic uncertainties) were derived as a function of η-cluster and E_T[56]. They are applied to simulation as an event weight to take into account the differences in lepton efficiencies between data and simulation. However, these scale factors are around 1 and their impact in the analysis is quite small.

The smearing or scaling of the reconstructed objects at the analysis level is a common procedure. They are used to match the object energy in simulation to the one in data, to match the object energy to a known quantity or to implement an

uncertainty in the analysis. The smearing process consists in changing the object energy distribution using random numbers from a given distribution, usually a Gaussian. In the case of electrons, the EM electron cluster energy in data was corrected by applying energy scales as a function of η, ϕ and E_T to match the Z boson peak mass. They were obtained from $Z \to ee$, $J/\Psi \to ee$ or $\frac{E}{p}$ studies using isolated electrons from $W \to e\nu$. The cluster energy was smeared in Monte Carlo samples to match the energy resolution in data and to adjust the width of the Z peak.

3.5 Muons

3.5.1 Muon reconstruction and identification

The reconstruction of muons relies on informations coming from the MS, the ID and the calorimeters. Its aim is to reconstruct muons with high efficiency in a wide p_{T} spectrum (from few GeV to the TeV scale). In ATLAS four muon categories are used, each optimized for a different need:

- **Standalone Muons:** to build Stand-Alone (SA) muons only the informations collected by the MS are used: the hits left in the three MDT stations are combined to form three segments, and the three segments are used as input to a fitter to reconstruct the track. Once the track is reconstructed it is back-extrapolated to the interaction point taking into account the energy loss in the calorimeter by means of a parametrization. The SA muons are reconstructed in a wide η region ($|\eta| < 2.7$), but suffer of the spectrometer inefficient regions at $\eta = 0$ and $\eta \sim 1.2$.

- **Combined Muons:** Combined (CB) muons are obtained combining SA muons with an ID tracks in terms of η, ϕ and p_{T}. The CB muons have the best resolution on the muons parameters: the p_{T} resolutions benefits of the long lever arm and of the precision of the two independent measurements used as input (SA and ID tracks), while the vertex parameters are provided by the ID track. The reconstruction efficiency of CB muons is $\sim 92\%$ at the plateau.

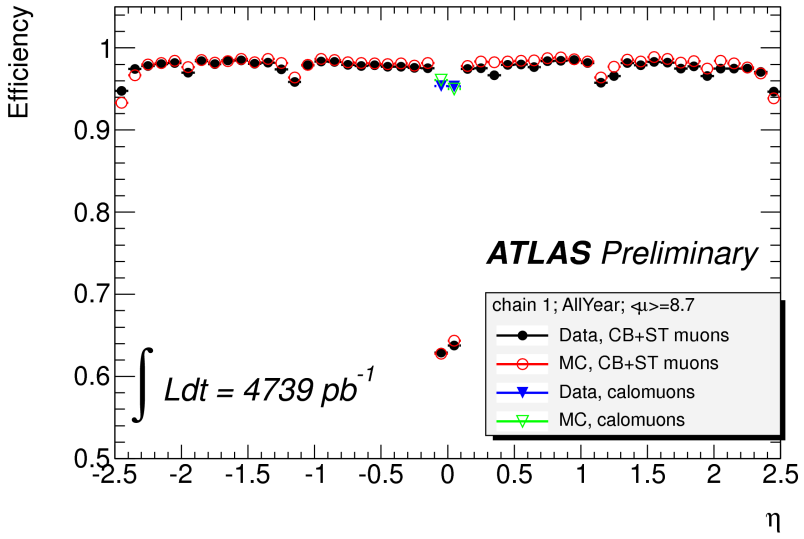
- **Tagged Muons:** The Tagged muon (ST) category is aimed at maximize the muon reconstruction efficiency in the low-p_{T} region: a low-p_{T} muon may not reach the spectrometer middle station because of magnetic field or might not penetrate the outermost calorimeter layers, and thus might not be reconstructed by the SA or CB algorithms. This muon category is build with an inside-out method which extrapolates an ID track to the entrance of the MS, and look for nearby hits in the first layer of the muon chambers. The reconstruction efficiency for ST muons is $\sim 98\%$ with $p_T > 40$ GeV.

- **Calorimeter Muons:** The Calorimeter muons also are reconstructed using an inside-out algorithm: here the ID track is matched with a calorimetric deposit compatible with a muon signature. this category is mainly used to compensate the loss in efficiency for $\eta \approx 0$.

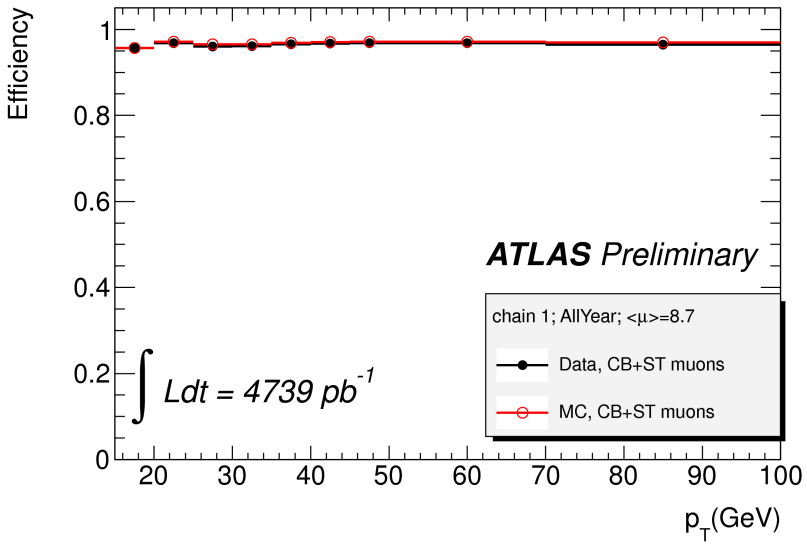
Loose
Middle-layer shower shapes: R_η, w_2
Hadronic leakage: R_{had1} (R_{had} for $0.8 < \|\eta\| < 1.37$)
Loose++
Shower shapes: $R_\eta, R_{had1}(R_{had}), w_2, E_{ratio}, w_{s,tot}$
Track quality
$\|\Delta\eta\| < 0.015$
Medium
Pass *Loose* selection
Strip-layer shower shapes: $w_{s,tot}$, E_{ratio}
Track quality
$\|\Delta\eta\| < 0.01$
$\|d_0\| < 5mm$
Medium++
Shower shapes as Loose++, but at tighter values
Track quality
$\|\Delta\eta\| < 0.005$
$N_{BL} \geq 1$ for $\|\eta\| < 2.01$
$N_{Pix} > 1$ for $\|\eta\| > 2.01$
Loose TRT HT fraction cuts
$\|d_0\| < 5mm$
Tight
Pass Medium selection $\|\Delta\eta\| < 0.005$
$\|d_0\| < 1mm$
Track matching: $\|\Delta\phi\|$ and E/p
High TRT HT fraction
$N_{BL} \geq 1$
Pass conversion bit
Tight++
Shower shapes as Medium++, but at tighter values Track quality $\|\Delta\eta\| < 0.005$
$N_{BL} \geq 1$ for all η
$N_{Pix} > 1$ for $\|\eta\| > 2.01$
Tighter TRT HT fraction cuts
$\|d_0\| < 1mm$
E/p requirement
$\|\Delta\phi\|$ requirement
Conversion bit

Table 3.1 Summary of the variables used in the Loose, Loose++, Medium, Medium++, Tight and Tight++ operating points[57].

In ATLAS, muons can be reconstructed using two independent algorithms: STACO [87] and MUID [88], and each of them provides algorithms to reconstruct all the four muon categories. The two families use different approaches to reconstruct the tracks starting from the hit in the detectors, and it has been proven that they have very similar performances [87, 88].



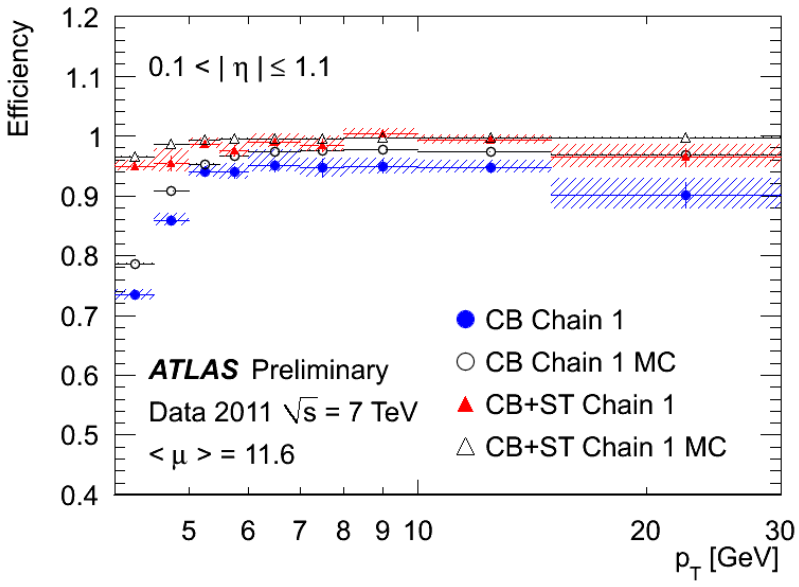
(a)



(b)

Figure 3.5 Muon reconstruction efficiency as a function of η (a) and p_{T} (b) in 2011 pp collisions for the STACO. family

Anyway the default algorithm used for physics analyses is STACO, some of its performances are shown in Figure 3.6. As can be seen from Figure 3.6, there is a good agreement between 2011 data and Monte Carlo in both the p_{T} and the η spectra. The η inefficiencies are due to barrel→End-Cap transition. The feet affect only the ϕ performance (not showed here); while the $\eta = 0$ region is where the two halves of the ATLAS muon spectrometer (lying in the $|\eta| > 0$ and $|\eta| < 0$ regions) come close. In this region the coverage of the muon spectrometer is not optimal (See



(a)

Figure 3.6 Muon reconstruction efficiency as a function of low-p_{T} in 2011 pp collisions for the STACO. family

Figure 3.5).

3.5.2 Muon scale factors and energy corrections

Trigger, reconstruction and identification (including isolation) efficiency scale factors derived using $Z \to \mu\mu$ events, are applied to Monte Carlo as a function of η, ϕ and the data period. Energy scale shifting and resolution smearing to correct the reconstructed muon momentum in MC are applied to match the energy scale and resolution measured in data $Z \to \mu\mu$ events. As tracks can be build from the ID and MS, there are separate corrections on the tracks in each subdetectors.

3.6 Jets

At high energy *pp* collisions the presence of partons is overwhelming. Due to colour confinement the partons hadronize. While the resulting bunch of particles passes through the ATLAS detector, they produce tracks in the ID and energy deposits inside the calorimeters. These detector signals allow the reconstruction of track jets (reconstructed using track information) and calorimeter jets (reconstructed using calorimeter information). The different analyses presented in this thesis make use of calorimeter jets. This section will then focus in explaining the jet reconstruction process for calorimeter jets only. It consists in three steps: the definition of calorimeter signals, the use of a jet reconstruction algorithm to group the calorimeter signals and finally the jet calibration which corrects the jet energy and momentum for the effects of ATLAS calorimeters non-compensation, dead material, leakage, out of cone and other thresholds effects.

3.6.1 Jet reconstruction

Inputs to calorimeter jet reconstruction

In a first step, calorimeter cells are combined and the resulting clusters will be used as inputs for the jet reconstruction algorithm. ATLAS provides two different cells clustering algorithms and therefore two types of inputs to jet reconstruction:

- **Topological clusters or topoclusters** are a dynamically formed combination of cells around seed cells that exceed a given signal-to-noise ratio threshold. The seeds are defined to be the cells with $|E_{cell}/\sigma_{cell}^{noise}| > 4$, where E_{cell} is the cell energy and σ_{cell}^{noise} is the RMS of the cell noise distribution. Subsequently, their neighbouring cells are included if their signal-to-noise ratio exceeds a second threshold $|E_{cell}/\sigma_{cell}^{noise}| > 2$[58].

  Finally all cells neighbouring the formed topoclusters are added to the topocluster. Topoclusters are defined as massless. Their energy is obtained summing up the energy of all the cells included. Their direction is calculated from weighted averages of the pseudorapidity and azimuthal angles of the constituent cells relative to the nominal ATLAS coordinate system. The weight used is the absolute cell energy. Because of calorimeter noise fluctuations cluster can have a negative energy. Negative energy clusters are rejected entirely from the jet reconstruction since they do not have physical meaning.

- **Noise Suppressed Towers** are constructed by projecting calorimeter cells onto a grid with tower bin size $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$. When cells larger than 0.1 exist, like in the third layer of the tile calorimeter, they are splitted between towers, and so is their energy, in a proportional manner. The towers are built using only cells belonging to topological clusters. Therefore, the same noise suppression is used in both cases. Towers are also defined to be massless and their energy and direction are calculated in the same way as for topoclusters. Negative energy towers are rejected entirely from the jet reconstruction[58]. The difference between towers and clusters is illustrated in Figure 3.7.

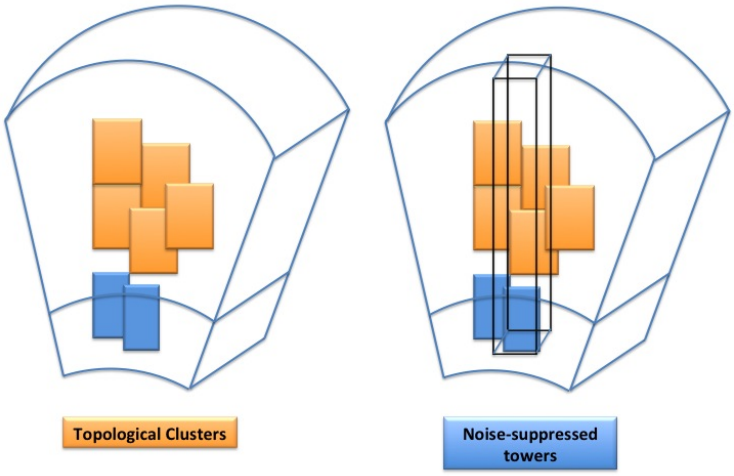


Figure 3.7 Representation of how cells may build a topological cluster (left) and how noise suppressed towers are built from those cells (right).

Jet reconstruction algorithms

Jet reconstruction algorithms allow to associate the energy deposits in the calorimeters to a jet. A good jet algorithm should give a stable and precise description of QCD interactions during the pp collision and therefore has to fulfill certain conditions:

- Collinear safety, which means that the splitting of one particle into two collinear particles has no effect on the reconstruction.
- Infrared safety, which means that the presence of additional soft particles between jet components does not affect the jet reconstruction.
- Effects of resolution and other detector effects (e.g. noise) should affect the jet reconstruction as little as possible.
- Invariance under Lorentz boosts along z coordinate.
- Minimum computer resources used.

The jet reconstruction algorithm used in this analysis is called the anti-k_T algorithm[59], which is a sequential recombination algorithm. Sequential recombination algorithms take topoclusters or towers as input and combines them to form jets according to a distance parameter defined below. For all inputs i, and pairs ij two different distances are defined:

$$d_{ij} = min(p_{T,i}^{2p}, p_{T,j}^{2p}) \frac{\Delta R_{ij}^2}{R^2} \tag{3.1}$$

$$d_i = p_{T,i}^{2p}, \tag{3.2}$$

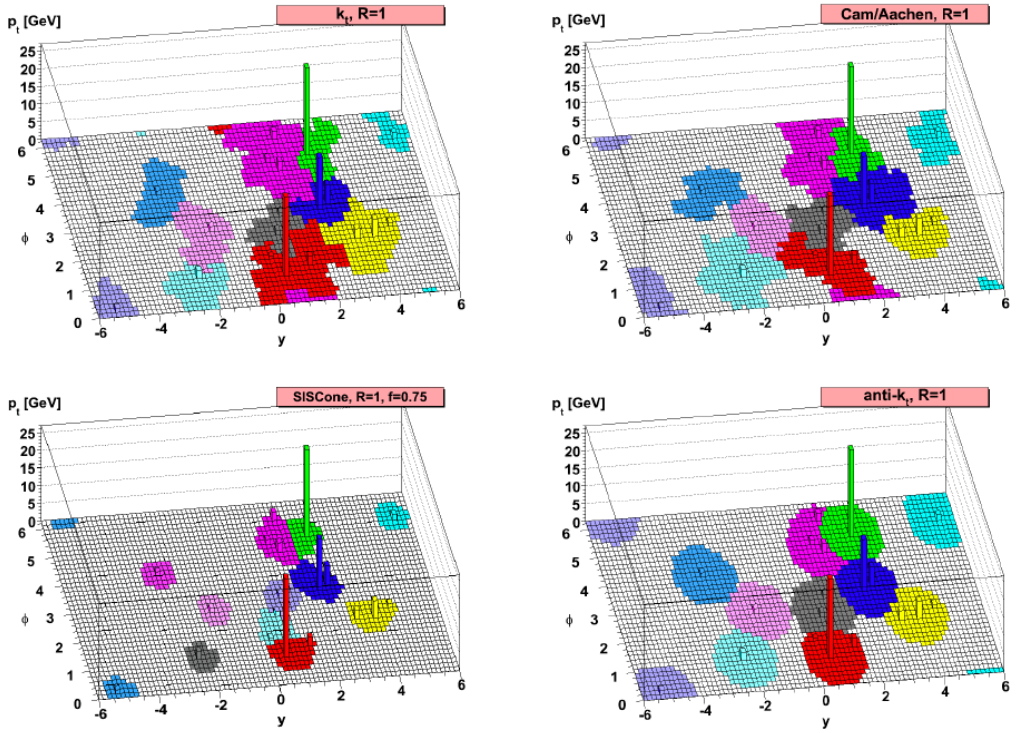


Figure 3.8 Shape output of the different reconstruction jet algorithms for the same event.

where p_T is the transverse momentum of the input i,

$$\Delta R_{ij} = \sqrt{(y_i - y_j)^2 + (\phi_i - \phi_j)^2} \tag{3.3}$$

is the distance between a pair of inputs in the $y - \phi$ space, y is the rapidity and R and p are parameters of the algorithm. d_{ij} represent the distance between a pair of inputs i and j, while d_i the distance between the input i and the beam axis in the momentum space. The algorithm calculates $min(d_i, d_{ij})$. If $min(d_i, d_{ij}) = d_i$, the input i is said to form a jet and is removed from the list of inputs. If $min(d_i, d_{ij}) = d_{ij}$, the inputs i and j are combined into one single input using the E-scheme (sum of four-momentum of each input). The combined input is put into the list of possible inputs, while i and j are removed. The algorithm proceeds until no inputs are left, which means that all inputs in the event will end in a jet. The parameter p defines the kind of algorithm: (Figure 3.8)

- p = 1 : k_T algorithm[60]
- p = 0 : Cambridge/Aachen algorithm[61]
- p = -1: anti-k_T algorithm[59]

While R characterizes the size of the jet in the $y-\phi$ space. The anti-k_T algorithm works in the inverse transverse momentum space and has three main advantages.

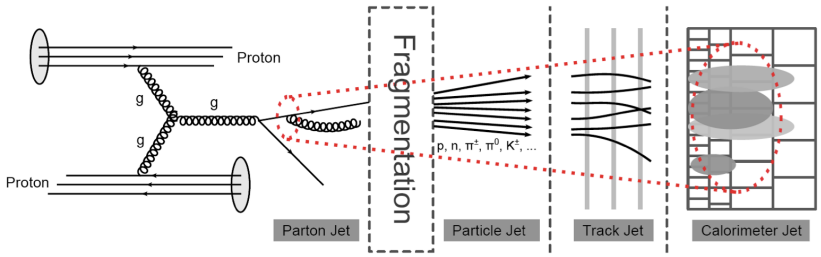


Figure 3.9 Schematic view of the different types of jets. Colorful particles, produced at the interaction point, create a bunch of colorless particles due to fragmentation. These particles will produce detector signals in the Inner Detector and Calorimeter, which can be reconstructed as track and calorimeter jets.

First, it clusters nearby particles, ensuring infrared safety. Second, soft inputs prefer to cluster with hard inputs instead of clustering with other soft particles. Third, the anti-k_T algorithm is seedless and all hard inputs within $\Delta R_{ij} < R$ will be combined into one jet, ensuring the collinear safety. In this analysis, the ratio parameter $R = 0.4$ was used.

The topoclusters and towers are defined massless. However, their distribution inside the reconstructed jet leads the jet to have a given mass. In Monte Carlo simulations track jets and calorimeter jets are reconstructed as in real data. In addition, two other kinds of jets can be defined in simulation:

- **Parton jet** refers to the parton at the matrix-element level which causes the particle shower due to the fragmentation process.
- **Particle or truth jet** which is reconstructed from stable particles (particles with a lifetime greater than 10 *ps*) produced by the hadronization models of the different Monte Carlo generators, before any detector simulation. This mainly includes electrons, photons, pions, kaons, protons and neutrons and their antiparticles. Neutrinos and muons are not included, since they do not leave any significant signal in the calorimeter.

A schematic view of the different types of jets is shown in Figure 3.9. The jet reconstruction efficiencies were determined from data with a tag-and-probe method, using track jets, where the efficiency was defined as the fraction of probe track jets matching a corresponding calorimeter jet[62]. The difference between data and simulation is found to be small and within the uncertainties. Therefore, no scale factors were needed.

3.6.2 Jet calibration

The energy of the reconstructed jets does not correspond to the initial energy carried by the particles. Reconstructed jets need therefore to be calibrated to the correct

energy scale. In general, the reference scale in the jet calibration process is given by the truth jets. Jets are initially reconstructed at the electromagnetic (EM) scale, which is the basic calorimeter signal scale for the ATLAS calorimeters. This means that the calorimeter signals are calibrated to properly reproduce the energy lost in the calorimeter by an electron, if the energy deposit came from an electron. The EM scale was obtained using test-beam measurements for electrons in the barrel[53, 54, 63] and the endcap calorimeters[65]. It has been validated using muons from test-beams and in cosmic-rays.
The energy scale of the electromagnetic calorimeters has been corrected using the invariant mass of $Z \rightarrow ee$ events[52]. This EM scale calibration provides a very good description for energy deposits produced by electrons and photons, but not for deposits from hadronic particles like protons, neutrons, pions or kaons. This difference results from the fact that the EM scale calibration does not account for various detector effects:

- Calorimeters non-compensation: partial measurement of the energy deposited by hadrons.
- Dead material: energy losses in inactive or non instrumented regions of the detector.
- Leakage: energy deposits from particles which are not totally contained in the calorimeters.
- Out of cone: loss of energy deposits from particles inside the particle jet that are not included in the reconstructed jet.
- Thresholds effects: signal losses due to inefficiencies in calorimeter clustering and jet reconstruction.

ATLAS has developed several calibration schemes with different levels of complexity and different sensitivity to systematic effects[58], which made them complementary in the way they contribute to the understanding of the jet energy scale measurement:

- **EM+JES calibration** is a Monte Carlo-derived jet calibration scheme used in ATLAS for the first analyses. EM+JES applies a simple jet-by-jet correction that restores the reconstructed jet energy to the particle jet energy. In this scheme each jet at the EM scale is scaled by a correction factor which is a function of the reconstructed jet energy and η. In addition to this energy correction a pile-up and a jet origin correction are also applied[58].
- **Global Sequential (GS) Calibration** is a Monte Carlo-derived jet calibration, which uses longitudinal and transverse properties of the jet structure sequentially to improve the resolution, while leaving the jet energy scale unchanged. In this scheme jets are found from clusters or towers, then the EM+JES calibration is applied and finally they are scaled by a jet-by-jet correction factor which depends on the jet p_T, η and several longitudinal and transverse jet properties.

- **Global Cell Energy-Density Weighting Calibration (GCW)** attempts to compensate for the different calorimeter response to hadrons and electromagnetic particles by weighting each jet constituent cell. The weights, which depend on the cell energy density and the calorimeter layer only, are determined by minimizing the energy fluctuations between the reconstructed and particle jets in Monte Carlo simulation. Jets are found from topoclusters or towers at the EM scale, then cells are weighted and a final jet energy scale correction is applied to ensure that good linearity response is achieved[66].
- **Local Cluster Weighting Calibration (LCW)** uses properties of topoclusters (such as their energy, depth in the calorimeter, cell energy density, fractional energy deposited in the calorimeter layer and energy measured around it) to calibrate them individually before applying jet reconstruction. These weights are determined from Monte Carlo simulations of charged and neutral pions. Similarly to the GCW scheme, a final correction of the jet energy is applied[66].

The EM+JES calibration is used for first physics analysis, due to its simplicity. The others calibration schemes are presently commissioned by ATLAS. The corrections applied by each calibration schemes as well as the inputs used have been validated using data from *pp* collision at a center-of-mass energy of $\sqrt{s} = 7$ TeV[58].

3.6.3 Jet selection

In ATLAS jets can be reconstructed with different quality criteria: *Looser*, *Loose*, *Medium* and *Tight* [89].

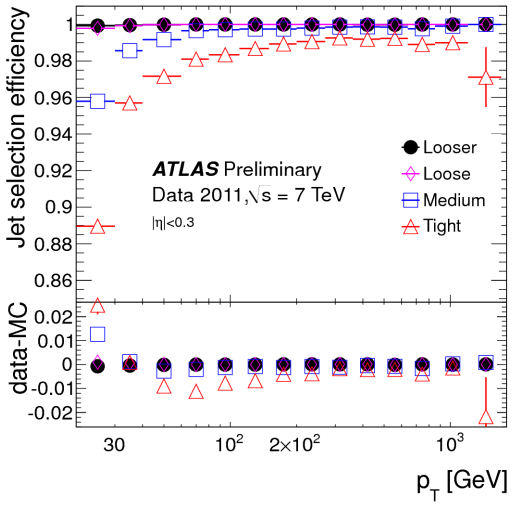
Since the noisy channels of the calorimeter and its electronics can lead to fake energy deposits not due to particles going through the calorimeter (which can be reconstructed as fake jets), many quality criteria on the features of the recorded pulse are applied in order to discriminate between real and fake jets candidates. The four jet categories differ for the cuts applied on the calorimetric variables of the signal.

The Looser selection was designed to provide an efficiency above 99.8% with a fake rejection as high as possible while the Tight selection was designed to provide a much higher fake rate jet rejection with an inefficiency not larger than a few percent. The two other sets of cuts correspond to intermediate fake rejections and jet selection efficiencies. Efficiencies of the four jet reconstruction categories are shown in Figure 3.10 as a function of the jet p_{T}.

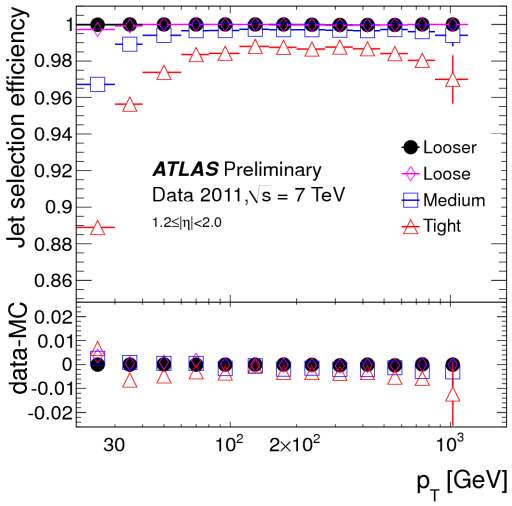
As explained above the jets in ATLAS are reconstructed using solely the calorimeter, however tracks reconstructed in the ID can be associated to a jet. Tracks are associated to jets using a simple geometrical matching criterion: the radial distance:

$$\Delta R = \sqrt{(\eta^{\mathrm{PV}}_{\mathrm{jet}} - \eta^{\mathrm{PV}}_{\mathrm{track}})^2 + (\phi^{\mathrm{PV}}_{\mathrm{jet}} - \phi^{\mathrm{PV}}_{\mathrm{track}})^2} \tag{3.4}$$

is calculated for each track, where $\eta^{\mathrm{PV}}_{\mathrm{jet}}$ and $\phi^{\mathrm{PV}}_{\mathrm{jet}}$ are the pseudorapidity and the azimuthal angle of the jet with respect to the primary vertex, and $\eta^{\mathrm{PV}}_{\mathrm{track}}$ and $\phi^{\mathrm{PV}}_{\mathrm{track}}$



(a)



(b)

Figure 3.10 Jet quality selection efficiency for anti-k_T jets with $R = 0.4$ as a function of p_{T} in two η example-ranges, for the four sets of selection criteria.

are the pseudorapidity and the azimuthal angle of the track at the perigee[2] with respect to the PV.

Any track for which the condition $\Delta R < 0.4$ is satisfied are considered as matching the jet, and therefore are associated to it.

3.6.4 Jet energy resolution

The procedure adopted to measure the jet energy resolution is explained in detail in [90]. It basically relies on the assumption that in events containing only two jets, the p_{T}'s of the two jets shall be balanced because of the momentum conservation in the transverse plane. Starting from this assumption the jet energy resolution can be measured studying the asymmetry observed between the jet p_{T}'s in such a configuration. To perform this measurement, jets in the same rapidity y region are chosen in order to minimize additional detector effects that may introduce secondary effects. The jet energy resolution is thus obtained in $p_{\mathrm{T}} \times \eta$ bins. Given the above, the fractional jet energy resolution can be parametrized as (following the parametrization of the energy resolution for the calorimeter):

$$\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} = \frac{N}{p_{\mathrm{T}}} \oplus \frac{S}{\sqrt{p_{\mathrm{T}}}} \oplus C \tag{3.5}$$

Where N, S, and C are the noise, stochastic and constant terms respectively. Once the measurement according to the method mentioned above is performed (and validated with Monte Carlo simulations), the distribution of the results can be built in each η bin and a fit with the functional form in equation 3.5 can be done. Some results of such a measurement is shown in Figure 3.11 for a specific rapidity bin.

In the rapidity bin $0.0 < |y| < 0.8$ shown in Figure 3.5, a $\sigma(p_{\mathrm{T}})/p_{\mathrm{T}}$ of about 15% is reached for jets having $p_{\mathrm{T}} = 40$ GeV, while at $p_{\mathrm{T}} = 500$ GeV $\sigma(p_{\mathrm{T}})/p_{\mathrm{T}} \sim 7\%$. Once a measurement of the jet energy resolution is obtained, it is possible to link any measured jet falling in a given $\eta - p_{\mathrm{T}}$ region to its corresponding resolution. This correspondence has been used as an ingredient of the jet pairing algorithm described in section 5.3.3.

[2]The perigee of a track is defined as point of closest approach to the beam axis

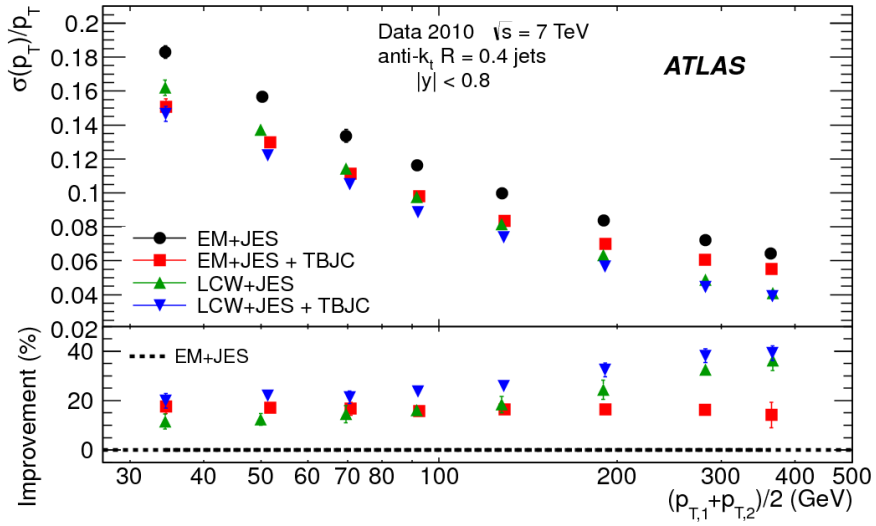


Figure 3.11 Fractional jet p_T resolutions, measured in data for anti-kt jets with $R = 0.4$ and for four jet calibration schemes: EM+JES, EM+JES+TBJC, LCW+JES and LCW+JES+TBJC. The lower panel of the figure shows the relative improvement for the EM+JES+TBJC, LCW+JES and LCW+JES+TBJC calibrations with respect to the EM+JES jet calibration scheme, used as reference (dotted line). The errors shown are only statistical.

3.7 $b-$tagging

3.7.1 $b-$tagging algorithms

The B hadron formed by the bottom quark has a relatively long lifetime of about $1 \times 10^{-12}s$ and can travel around $3mm$ before decaying. The identification of $b-$jets is very important for example in the discrimination of top quark analysis backgrounds with only light jets in the final state. $b-$tagging algorithms exploit the fact that a certain number of tracks point to a secondary vertex instead of pointing to the reconstructed primary vertices as shown in Figure 3.12 and that impact parameters of these tracks are large. Thus $b-$tagging relies on the ID track reconstruction and for such reason it can be only applied to jets with $|\eta| < 2.5$. Moreover ID tracks are required to pass some quality criteria that depend on the different $b-$tagging algorithms.

Various $b-$tagging algorithms can be defined, based on these discriminating variables, on secondary vertex properties and on the presence of leptons within b-jets, and for each jet they usually give as output a weight reflecting the probability that the input jet originates from a $b-$quark. There are basically three kinds of $b-$tagging algorithms[67] The one used in this analysis is based on the measurement of the *impact parameter* (see Figure 3.12), and on the reconstruction of the secondary vertex.

The transverse impact parameter d_0 is the distance in the transverse plane $x - y$ between the point of closest approach of a track and primary vertex, while the

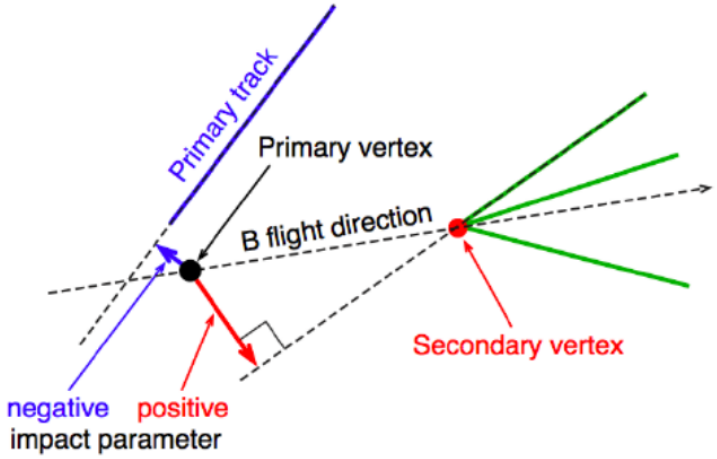


Figure 3.12 Schematic representation of a b-hadron decay and definition of the impact parameter.

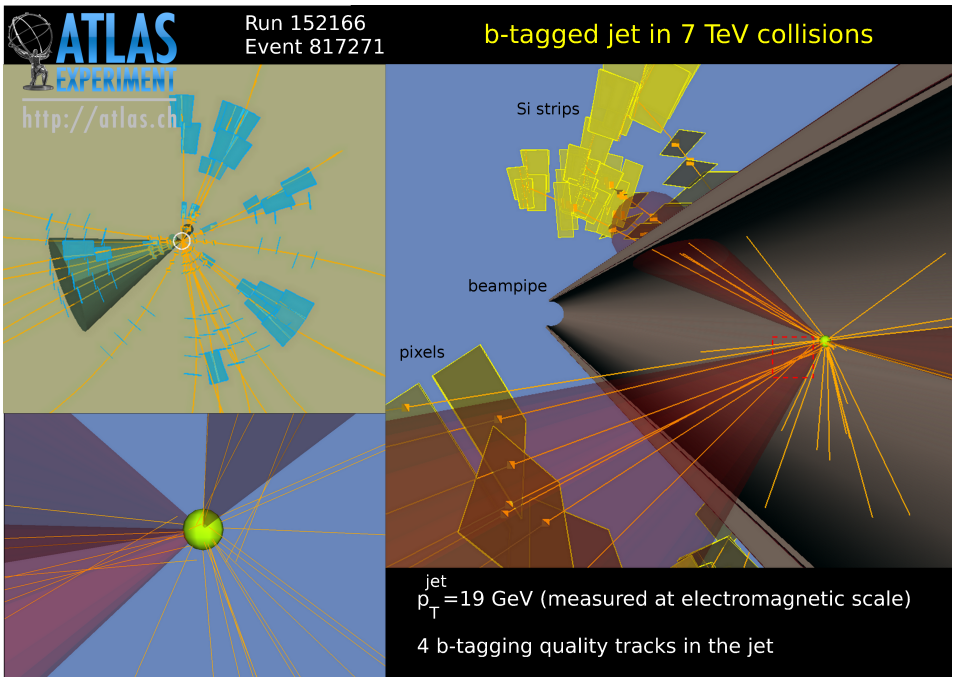
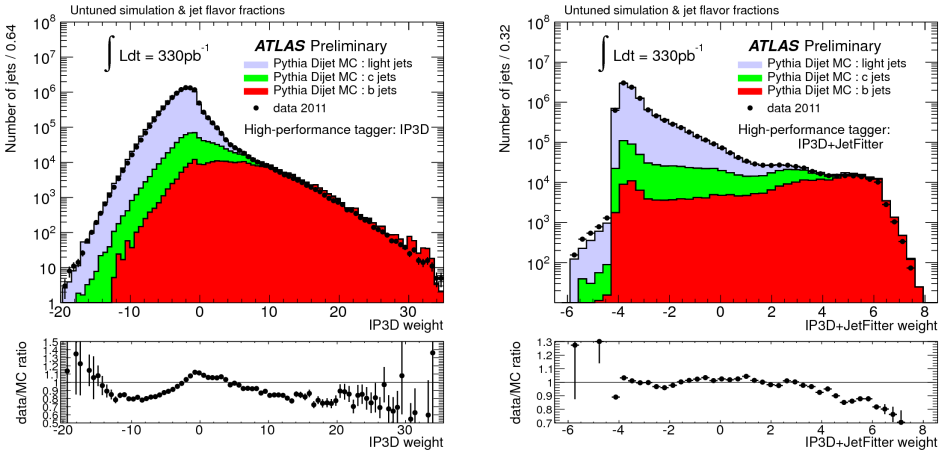


Figure 3.13 Real collision view of a $b-$jet candidate. The event display shows a $b-$jet candidate reconstructed with the anti$-k_T$ algorithm, where the primary vertex is shown in the yellow circle and the secondary vertex in the dashed red box.

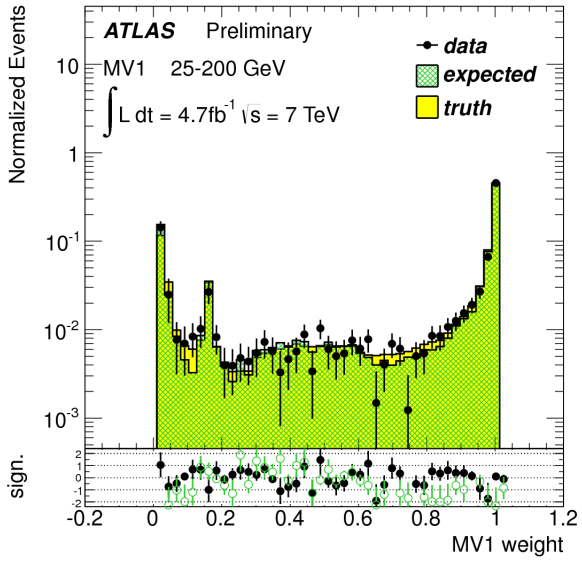
longitudinal impact parameter is the $z-$coordinate of this point (z_0). The $b-$tagging algorithm used in these analyses is called *MV1* (multiVariate tagger) [92]: it is based on a neural network, and takes as input the output weights of three simpler tagging algorithms:

- IP3D [91], based on the d_0 significance

- SV1 [91], based on the reconstruction of secondary vertexes
- JetFitterCombNN [91], which performs a fit on the flight direction of the $b-$hadrons and then combines the result with the output weights of IP3D and SV1.



(a) (b)



(c)

Figure 3.14 Data-MC comparison of the output of the input taggers: IP3D 3.14(a), JetFitterCombNN 3.14(b) (also known as *JetFitter+IP3D*). The data-Monte Carlo comparison for the output of the MV1 tagger is also shown3.14(c).

Figures 3.14(a)-3.14(b) show the data-Monte Carlo comparison for the three taggers used as input to the MV1 algorithm. The output of the MV1 tagger is a continuous value w_{MV1}, and it is possible to choose a threshold value $\bar{w}$ to tag a jet as a $b-$jet: if $w_{MV1} > \bar{w}$ then the jet will be considered a $b-$jet, otherwise it will be

considered a light-flavor jet. The data-Monte Carlo comparison on the output weight of the MV1 algorithm is shown in Figure 3.14(c). As can be seen the data-Monte Carlo agreement is good for the considered taggers. In particular the output weight of the MV1 tagger clusters to 0 for light jets, while assumes values near to 1 for $b-$jets. Moreover in figure 3.14(c) it can be seen that the output weight of the MV1 tagger has a peak at $w \sim 0.15$. It is due to jets coming from the hadronization of the $c-$quark. The choice of the value of $\bar{w}$ depends on the desired $b-$tagging efficiency and on the desired mistag rate one wants to have in the analysis as shown in Figures 3.15 and 5.12.

3.7.2 $b-$tagging scale factors

The performance of the $b-$tagging algorithm is given by the efficiency with which a $b-$jets is tagged as such, by the $c-$tag efficiency, which is the equivalent quantity for jets originating from $c-$quarks, the τ-tag, the light jet tag efficiency and by the corresponding mistag rates. They are not the same in data and simulation, so we need to apply scale factors to correct the simulation[68, 69, 70, 71]. In release 16, for tau jets the light jets efficiency and mistag SFs were used, but with twice the uncertainty, since the tau jets have few tracks and are more similar to light jets. For release 17, additional SFs for $\tau-$leptons were derived. $b-$tagging scale factors have values close to 1, but their systematic uncertainties are large.

3.8 Missing transverse energy

The missing transverse energy, $E_{\mathrm{T}}^{\mathrm{miss}}$, in a collider experiment is the energy imbalance in the transverse plane, where the energy conservation is expected. The physical source of such an imbalance typically is the presence of unseen particles such as neutrinos which go through all the detector without leaving any signal, and it is measured thanks to the tightness of the calorimetric system.

In addition many detector-related effect (such as mismeasurements of energy) can give rise to $E_{\mathrm{T}}^{\mathrm{miss}}$. The $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction algorithm starts from all the calorimetric cells belonging to topological clusters (see section 3.10) in the $|\eta| < 4.9$ range, considering their energy and also their position in θ and ϕ, as shown in equation 3.9.

The final $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation is defined as:

$$E_{\mathrm{T}}^{\mathrm{miss}} = \sqrt{(E_x^{\mathrm{miss}})^2 + (E_y^{\mathrm{miss}})^2} \tag{3.6}$$

where the $E_{x(y)}^{\mathrm{miss}}$ contain contribution from both calorimetric energy deposits and corrections for the muons in the event in each transverse direction x and y.

$$E_{x(y)}^{\mathrm{miss}} = E_{x(y)}^{\mathrm{miss,\ Calo}} + E_{x(y)}^{\mathrm{miss,\ Muon}} \tag{3.7}$$

where

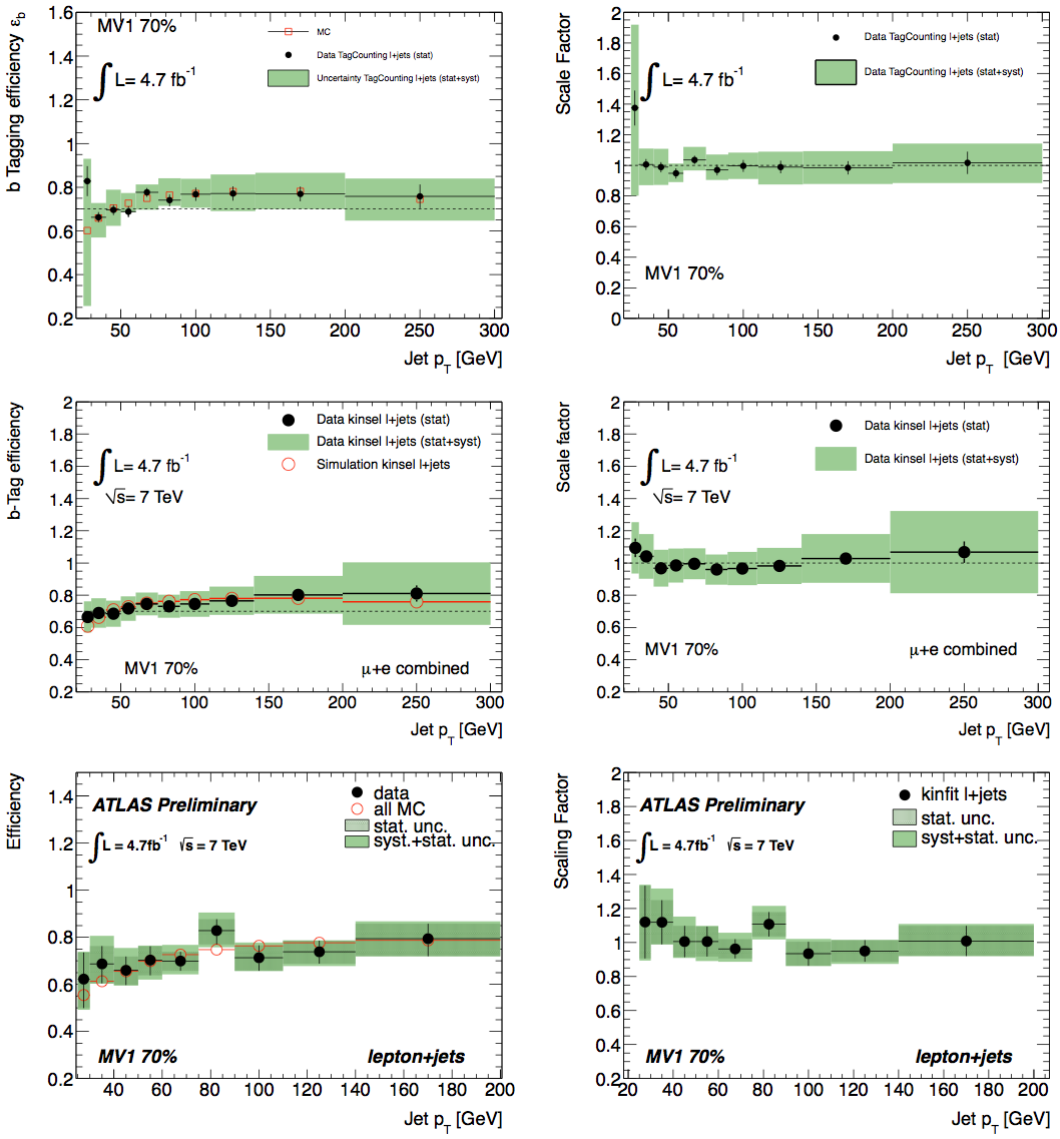
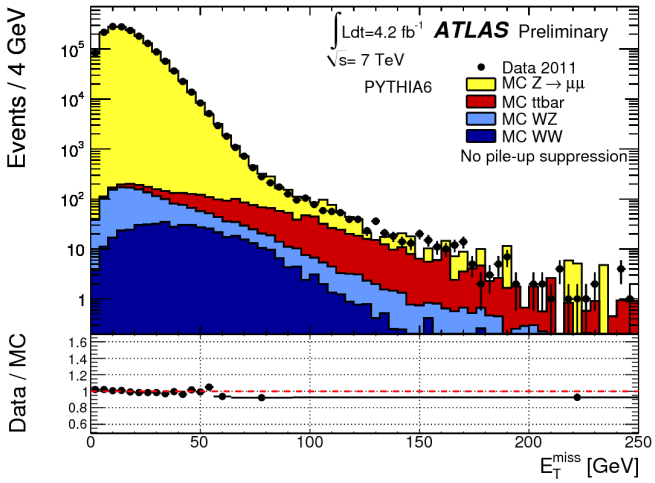


Figure 3.15 b-tagging efficiency estimated from data compared with true efficiency for sum of the Monte Carlo (left) and resulting scaling factors (right) together with statistical(bars) and systematic (band) uncertainties for the $MV1$ tagger at the working point corresponding to 70% efficiency for the single-lepton $t\bar{t}$ samples.

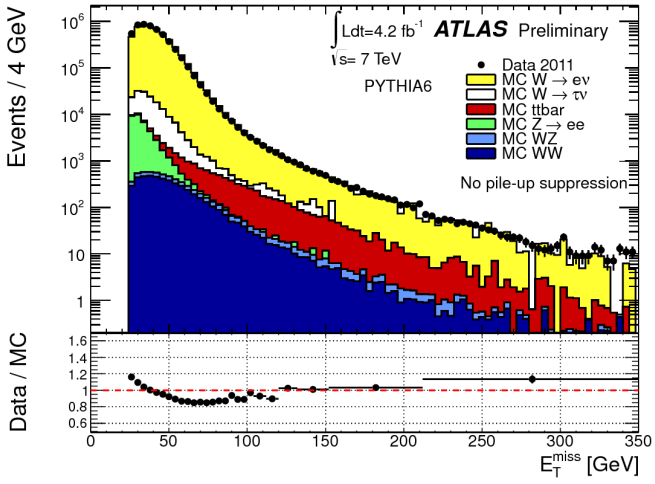
$$E_x^{\text{miss, Calo}} = -\sum_{i=1}^{N_{\text{cell}}} E_i \sin\theta_i \cos\phi_i \tag{3.8}$$

$$E_y^{\text{miss, Calo}} = -\sum_{i=1}^{N_{\text{cell}}} E_i \sin\theta_i \sin\phi_i \tag{3.9}$$

And the $E_{x(y)}^{\text{miss, Muon}}$ takes into account the energy muon energy deposit as it goes through the calorimetric system. The $E_{x(y)}^{\text{miss, Calo}}$ terms contain all the energy deposits



(a)



(b)

Figure 3.16 Distribution of $E_{\mathrm{T}}^{\mathrm{miss}}$ measured in $Z \to \mu^+\mu^-$ (a) and $W \to e\nu$ (b) events for data and Monte Carlo.

in the calorimeter: all the energy deposits associated to reconstructed physics objects (electrons, photons, taus, jets) are considered as well as those that are not associate to any reconstructed object. This last contribution may suffer of contamination from noisy channels, but this is avoided by means of quality requirement on any energy deposit contributing to the $E_{\mathrm{T}}^{\mathrm{miss}}$ calculation [93].

Figure 3.16 shows the reconstruction of $E_{\mathrm{T}}^{\mathrm{miss}}$ in $Z \to \mu^+\mu^-$ (Figure 3.16(a)) and $W \to e\nu$ (Figure 3.16(b)) for both data and Monte Carlo.

Figure 3.17 the $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution as a function of the number of reconstructed primary vertexes in each event. This shows how pileup affects calorimetric measurements and the effect of pileup suppression methods.

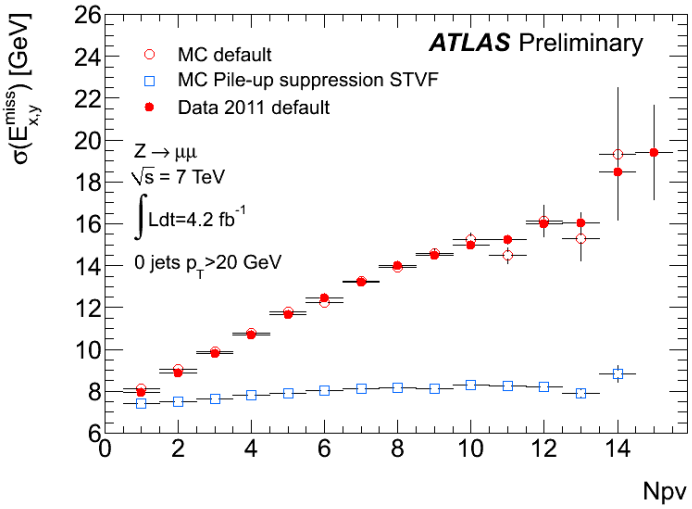


Figure 3.17 $E_{\mathrm{T}}^{\mathrm{miss}}$ resolution as a function of the number of reconstructed primary vertexes (i.e. as a function of the pileup). The effect of the pileup suppression is shown too.

As an example of how looks this kind objects inside ATLAS, the Figure 3.18 shows a high mass central dijet event collected by the end of April 2012: two central high$-p_T$ jets have an invariant mass of 3.65 TeV and the highest p_T jet has p_T of 1.72 TeV. A track p_T cut of 0.5 GeV has been applied for the display.

- 1st jet (ordered by p_T): p_T = 1.72 TeV, Îů = -0.04, ÏĘ = -2.68
- 2nd jet: p_T = 1.50 TeV, $\eta = 0.64$, $\phi = 1.70$
- 3rd jet: p_T = 0.22 TeV, $\eta = 0.28$, $\phi = -2.13$
- Missing E_T = 29 GeV, $\phi = 0.50$
- Sum E_T = 3.91 TeV

Event collected on 10 April 2012.

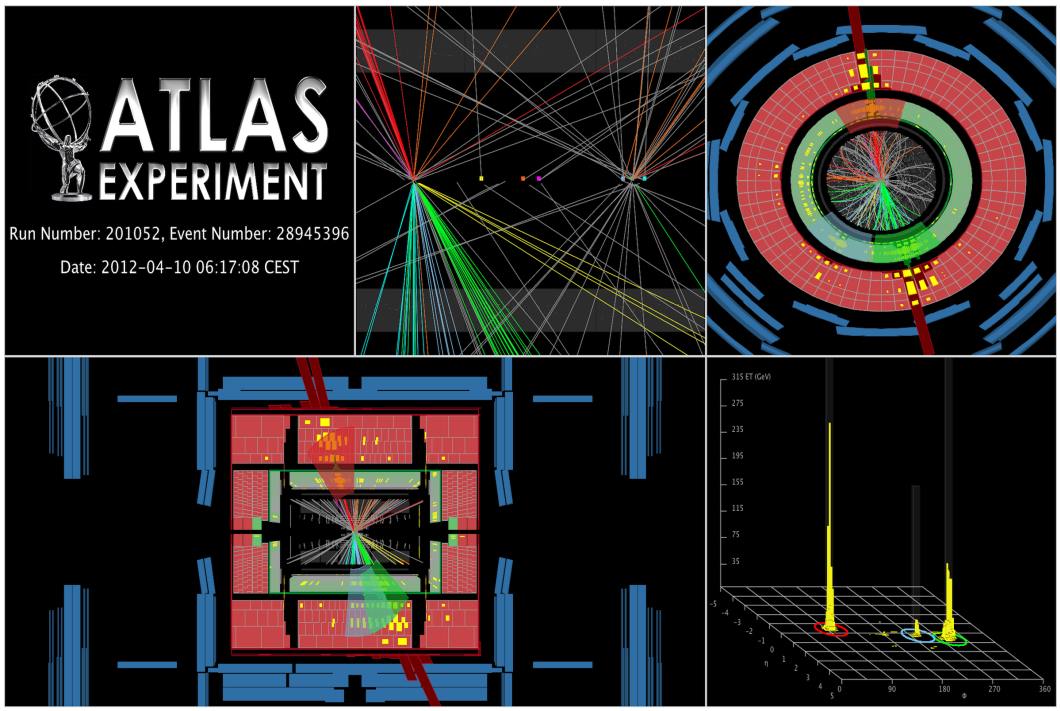


Figure 3.18 High mass central dijet event collected by the end of April 2012.

Chapter 4

The Standard Model Higgs search in ATLAS

This chapter describes the most updated ATLAS results for the SM Higgs boson search. All the analysis have been divided in "Low Mass" and "High Mass" searches, where the Low Mass range can be defined between 110 GeV and $\approx 2M_Z$, and the High Mass range can be defined between 200 GeV and 600 GeV. Some searches are designed to fully exploit the features of the production modes (see Section 1.3.2) $pp \to H$ (gluon gluon fusion), $pp \to qqH$ (vector boson fusion) and $pp \to VH$ with $V = W^{\pm}$ or Z (associate production with a gauge boson). The branching ratios of the different decay's modes of the SM Higgs boson are shown in Figure 4.1 and 4.2.

4.1 Low Mass search

4.1.1 $H \to \gamma\gamma$ channel

This analysis is carried out for m_H hypotheses between 110 GeV and 150 GeV. Events are separated into nine independent categories of varying sensitivity. The categorization is based on the $|\eta|$ of each photon, whether it was reconstructed as a converted or unconverted photon, and the momentum component of the diphoton system transverse to the diphoton thrust axis (p_{Tt}). The mass resolution is approximately 1.7% for $m_H \approx 120$ GeV[115, 116].

4.1.2 $H \to ZZ(^*) \to 4l$ channel

This analysis[117] is performed for m_H hypotheses in the 110 GeV to 600 GeV mass range. The main irreducible $ZZ(^*)$ background is estimated using a combination of Monte Carlo simulation and the observed data. The reducible Z+jets background, which mostly impacts the low four-lepton invariant mass region, is estimated from control regions in the data. The top-quark ($t\bar{t}$) background normalization is validated using a dedicated control sample. Four categories of events are defined by the lepton flavour combinations and the four-lepton invariant mass is used as a discriminating variable: $eeee$, $\mu\mu\mu\mu$, $\mu\mu ee$ and $ee\mu\mu$. The mass resolution is approximately 1.5% in the four-muon channel and 2% in the four-electron channel for $m_H \approx 120$ GeV.

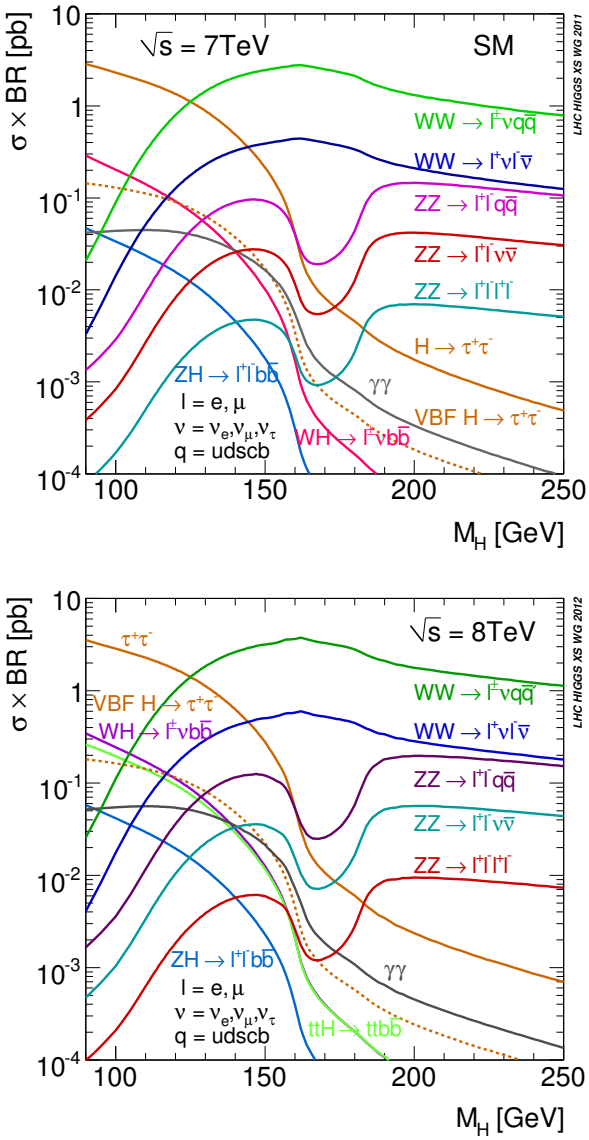


Figure 4.1 Standard Model Higgs boson decay branching ratios for LHC center of mass energy of $\sqrt{s} = 7$ TeV(a) and $\sqrt{s} = 8$ TeV(b) for Low Mass range.

4.1.3 $H \to ZZ(^*) \to l^{\pm}l^{\pm}q\bar{q}$ channel

This search is performed for m_H hypotheses ranging from 120 GeV to 180 GeV and *is the main subject of this dissertation.* The dominant background arises from $Drell - Yan$+Jets and in minor −but important− proportion from Z+jets production, which are treated together and have been estimated using side-bands of the dijet invariant mass distribution in data. To profit from the relatively large rate of b-jets from Z boson decays present in the signal compared to the rate of b-jets found in the Z+jets background, the analysis is divided into two categories. The first category contains events in which the two jets are b-tagged and the second uses events with less than two b-tags. Our analysis takes advantage of a highly

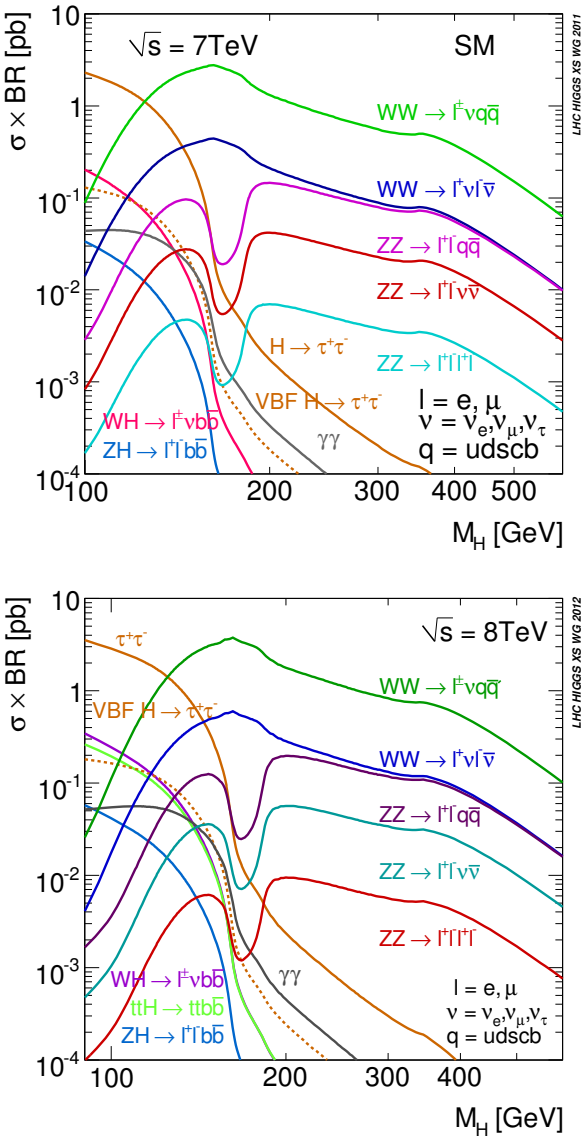


Figure 4.2 Standard Model Higgs boson decay branching ratios for LHC center of mass energy of $\sqrt{s} = 7$ TeV(a) and $\sqrt{s} = 8$ TeV(b) for Low and High Mass ranges.

efficient b-tagging algorithm[118] and the side-band to constrain the background yield. For first time, the use of a Kinematic Fitter tool (see Section 5.3.3) improve the resolution and selection efficiency of the signal. The invariant mass of the $l^{\pm}l^{\pm}qq$ system is used as a discriminating variable.

4.1.4 $H \to WW(^*) \to e\nu\mu\nu$ channel

The updated analysis[119] is performed for m_H values from 110 GeV up to 600 GeV, reason to be into the next section too, but looking into its strong value at Low Mass search, have been included here. Events with two leptons are classified by the number of associated jets (0, 1 or 2), where the two-jet category has selection criteria designed to enhance sensitivity to the VBF production process. The events are further divided by the flavours of the charged leptons, ee, $e\mu$ and $\mu\mu$ where the mixed mode ($e\mu$) has a much smaller background from the Drell-Yan process. As in the case of $H \to ZZ \to l^+l^-\nu\nu$, the samples are split according to the pile-up conditions and analysed separately. Each sub-channel uses the WW transverse mass distribution, except for the 2-jets category, which does not use a discriminating variable.

4.1.5 $H \to \tau^+\tau^-$ channel

The analyses are categorized by the decay modes of the two τ leptons, for m_H hypotheses ranging from 110 GeV to 150 GeV (the leptonically decaying τ leptons are denoted τ_{lep} and the hadronically decaying τ leptons are denoted τ_{had}). Most of these sub-channels are triggered using leptons, except for the fully hadronic channel $H \to \tau_{had}\tau_{had}$, which is triggered with specific double hadronic τ decay selections. All the searches using τ decay modes have a significant background from $Z \to \tau^+\tau^-$ decays, which are modelled using an embedding technique where $Z \to \mu^+\mu^-$ candidates selected in the data have the muons replaced by simulated τ decays. These embedded events are used to describe this background process.

Figure 4.3 shows a compendium of the different results for the Low Mass analysis into the ATLAS combination presented.

4.2 High Mass search

4.2.1 $H \to WW \to l\nu q\bar{q}'$ channel

This analysis is performed for m_H hypotheses ranging from 300 GeV to 600 GeV. A leptonically decaying W boson is tagged with an isolated lepton and missing transverse momentum (E_T^{miss}). Additionally, two jets with an invariant mass compatible with a second W boson[120] are required.

The W boson mass constraint allows the reconstruction of the Higgs boson candidate mass on an event-by-event basis by using a quadratic equation to solve for the component of the neutrino momentum along the beam axis. Events where this equation has imaginary solutions are discarded in order to reduce tails in the mass distribution[120].

The analysis searches for a "peak" in the reconstructed $l\nu q\bar{q}'$ mass distribution. The analysis is further divided by lepton flavour and by the number of additional jets (0, 1 or 2), where the two jet channel is optimized for the VBF production process.

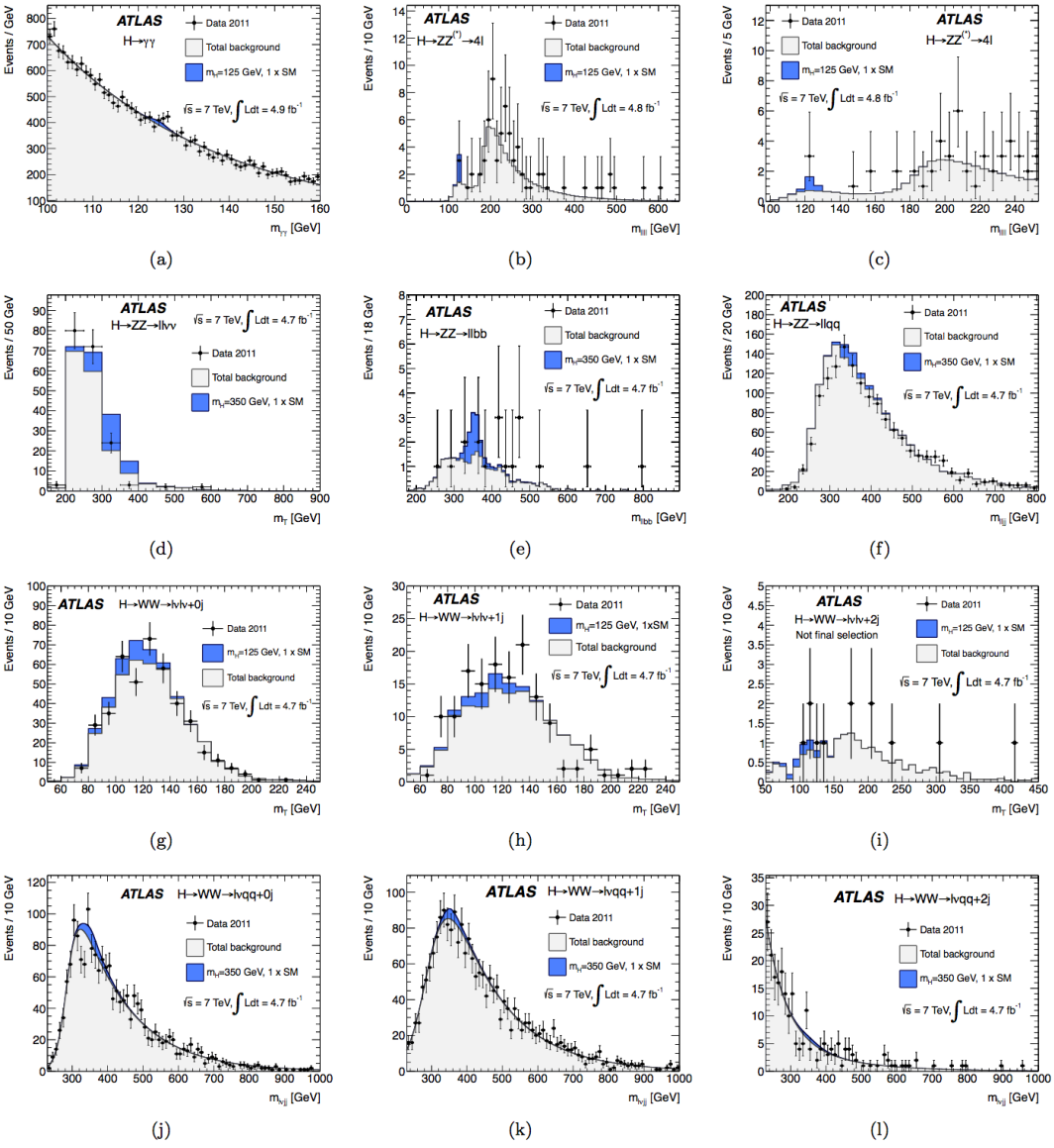


Figure 4.3 Invariant or transverse mass distributions for the selected candidate events, the total background and the signal expected in the following channels: (a) $H \to \gamma\gamma$, (b) $H \to ZZ(*) \to l^+l^-l^+l^-$ in the entire mass range, (c) $H \to ZZ(*) \to l^+l^-l^+l^-$ in the low mass range, (d) $H \to ZZ(*) \to l^+l^-\nu\bar{\nu}$, (e) $b-$tagged selection and (f) untagged selection for $H \to ZZ \to l^+l^-q\bar{q}$, (g) $H \to WW(*) \to l^+\nu l^-\bar{\nu} + 0-$jets, (h) $H \to WW(*) \to l^+\nu l^-\bar{\nu} + 1-$jet, (i) $H \to WW(*) \to l^+\nu l^-\bar{\nu} + 2-$jets, (j) $H \to WW(*) \to l^+\nu q\bar{q}'+0-$jets, (k) $H \to WW(*) \to l^+\nu q\bar{q}'+1-$jet and (l) $H \to WW(*) \to l^+\nu q\bar{q}' + 2-$jets. The $H \to WW(*) \to l^+\nu l^-\bar{\nu} + 2-$jets distribution is shown before the final selection requirements are applied.

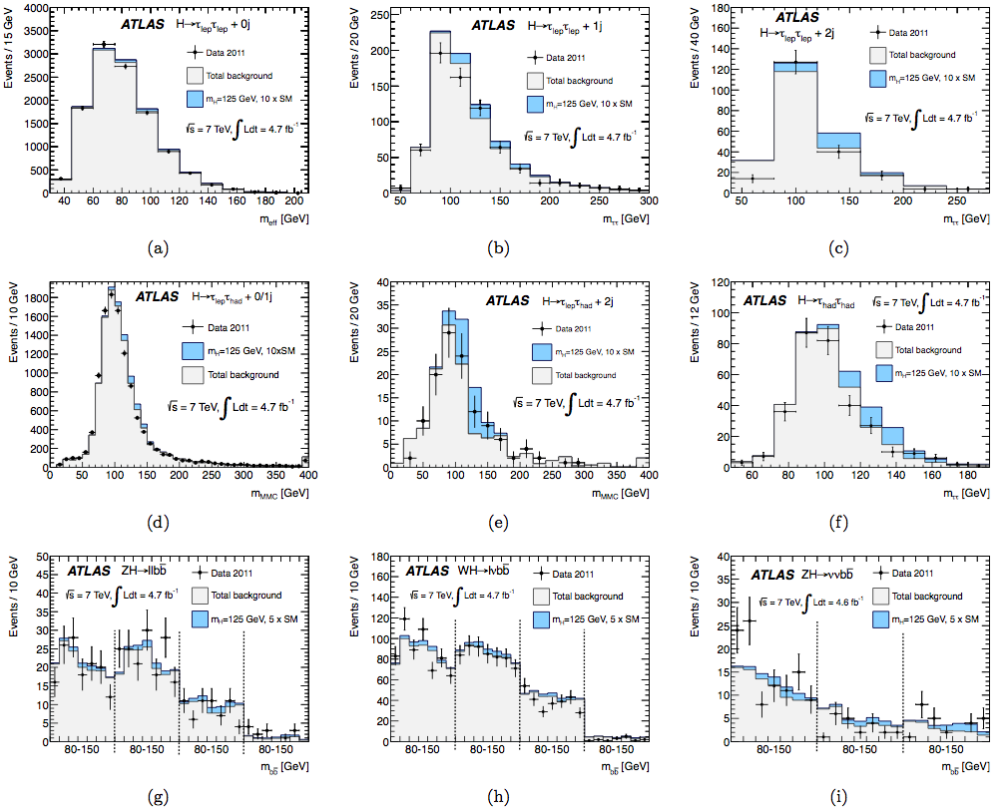


Figure 4.4 Invariant or transverse mass distributions for the selected candidate events, the total background and the signal expected in the following channels: (a) $H \to \tau_{lep}\tau_{lep} + 0$−jets, (b) $H \to \tau_{lep}\tau_{lep} + 1$-jet, (c) $H \to \tau_{lep}\tau_{lep} + 2$−jets, (d) $H \to \tau_{lep}\tau_{had} + 0$−jets and 1−jet, (e) $H \to \tau_{lep}\tau_{had} + 2$−jets, (f) $H \to \tau_{had}\tau_{had}$. The bb invariant mass for (g) the $ZH \to l^+l^-b\bar{b}$, (h) the $WH \to l\nu b\bar{b}$ and (i) the $ZH \to \nu\bar{\nu}b\bar{b}$ channels. The vertical dashed lines illustrate the separation between the mass spectra of the subcategories in p_T^Z, p_T^W, and E_T^{miss}, respectively. The signal distributions are lightly shaded where they have been scaled by a factor of five or 10 for illustration purposes.

4.2.2 $H \to ZZ \to l^{\pm}l^{\pm}q\bar{q}$ channel

This search is performed for m_H hypotheses ranging from 200 GeV to 600 GeV and is separated into search regions above and below $m_H = 300$ GeV, for which the event selections are independently optimized. The dominant background arises from Z+jets production, which is estimated using side-bands of the dijet invariant mass distribution in data. To profit from the relatively large rate of b-jets from Z boson decays present in the signal compared to the rate of b-jets found in the Z+jets background, the analysis is divided into two categories. The first category contains events in which the two jets are b-tagged and the second uses events with less than two b-tags. The analysis [19] takes advantage of a highly efficient b-tagging algorithm [27] and the side-band to constrain the background yield. Using the Z boson mass constraint improves the mass resolution of the system by more than a factor of two. The invariant mass of the $l^{\pm}l^{\pm}qq$ system is used as a discriminating variable.

4.2.3 $H \to ZZ \to l^{\pm}l^{\pm}\nu\nu$ channel

This analysis[121] is split into two regimes according to the level of pile-up, i.e. the average number of pp collisions per bunch crossing. The search is performed for m_H hypotheses ranging from 200 GeV to 600 GeV. The analysis is further categorized by the flavour of the leptons from the Z decay. The selection is optimized separately for Higgs boson masses above and below 280 GeV. The $l^{+}l^{\hat{a}\acute{L}\check{S}}$ invariant mass is required to be within 15 GeV of the Z boson mass. The inverted requirement is applied to same-flavour leptons in the $H \to WW(^{*}) \to l^{+}\nu l^{-}\nu$ channel to avoid overlap in the selection. The transverse mass (m_T), computed from the dilepton transverse momentum and the missing transverse momentum, is used as a discriminating variable.

As in the case of the Low Mass, Figure 4.4 shows a compendium of the different results for the Low Mass analysis into the ATLAS combination presented.

4.3 Combination

A recent and important result to be mentioned is the expected and observed limits from the individual channels entering in the combined search for the Standard Model Higgs boson in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector, shown in Figure 4.5. The combined 95% CL[122] exclusion limits are shown in Figure 4.6 as a function of m_H. Adding a final plot (Figure 4.7) where is resume the best-fit values of the strength parameter for three representative sample at Low Mass Higgs boson hypotheses.

Lets describe the procedure used for computing frequentist p-values uses for quantifying the agreement with the background-only hypothesis and for determining exclusion limits. The procedures are based on the profile likelihood ratio test statistic.

The parameter of interest is the overall signal strength factor μ, which acts as a scaling to the total rate of signal events. Often writen as $\mu = \sigma/\sigma_{SM}$, where σ_{SM} is the SM production cross-section.

Higgs Boson Decay	Subsequent Decay	Sub-Channels	$\int L\,dt$ [fb^{-1}]
		2011 $\sqrt{s}$ =7 TeV	
$H \to ZZ^{(*)}$	4ℓ	$\{4e, 2e2\mu, 2\mu2e, 4\mu\}$	4.6
$H \to \gamma\gamma$	–	10 categories $\{p_{Tt} \otimes \eta_\gamma \otimes \text{conversion}\} \oplus \{\text{2-jet VBF}\}$	4.8
$H \to \tau\tau$	$\tau_{lep}\tau_{lep}$	$\{e\mu\} \otimes \{\text{0-jet}\} \oplus \{\ell\ell\} \otimes \{\text{1-jet, 2-jet, } p_{T,\tau\tau} > 100 \text{ GeV}, VH\}$	4.6
	$\tau_{lep}\tau_{had}$	$\{e,\mu\} \otimes \{\text{0-jet, 1-jet, } p_{T,\tau\tau} > 100 \text{ GeV, 2-jet}\}$	4.6
	$\tau_{had}\tau_{had}$	{1-jet, 2-jet}	4.6
$VH \to Vbb$	$Z \to \nu\nu$	$E_T^{miss} \in \{120-160, 160-200, \geq 200 \text{ GeV}\} \otimes \{\text{2-jet, 3-jet}\}$	4.6
	$W \to \ell\nu$	$p_T^W \in \{< 50, 50-100, 100-150, 150-200, \geq 200 \text{ GeV}\}$	4.7
	$Z \to \ell\ell$	$p_T^Z \in \{< 50, 50-100, 100-150, 150-200, \geq 200 \text{ GeV}\}$	4.7
		2012 $\sqrt{s}$ =8 TeV	
$H \to ZZ^{(*)}$	4ℓ	$\{4e, 2e2\mu, 2\mu2e, 4\mu\}$	21
$H \to \gamma\gamma$	–	14 categories $\{p_{Tt} \otimes \eta_\gamma \otimes \text{conversion}\} \oplus \{\text{2-jet VBF}\} \oplus \{\ell\text{-tag, } E_T^{miss}\text{-tag, 2-jet VH}\}$	21
$H \to WW^{(*)}$	$e\nu\mu\nu$	$\{e\mu, \mu e\} \otimes \{\text{0-jet, 1-jet}\}$	13
$H \to \tau\tau$	$\tau_{lep}\tau_{lep}$	$\{\ell\ell\} \otimes \{\text{1-jet, 2-jet, } p_{T,\tau\tau} > 100 \text{ GeV}, VH\}$	13
	$\tau_{lep}\tau_{had}$	$\{e,\mu\} \otimes \{\text{0-jet, 1-jet, } p_{T,\tau\tau} > 100 \text{ GeV, 2-jet}\}$	13
	$\tau_{had}\tau_{had}$	{1-jet, 2-jet}	13
$VH \to Vbb$	$Z \to \nu\nu$	$E_T^{miss} \in \{120-160, 160-200, \geq 200 \text{ GeV}\} \otimes \{\text{2-jet, 3-jet}\}$	13
	$W \to \ell\nu$	$p_T^W \in \{< 50, 50-100, 100-150, 150-200, \geq 200 \text{ GeV}\}$	13
	$Z \to \ell\ell$	$p_T^Z \in \{< 50, 50-100, 100-150, 150-200, \geq 200 \text{ GeV}\}$	13

Table 4.1 Summary of the individual channels entering the combined results presented here. In channels sensitive to associated production of the Higgs boson, V indicates a W or Z boson. The symbols $\otimes$ and $\oplus$ represent direct products and sums over sets of selection requirements, respectively.

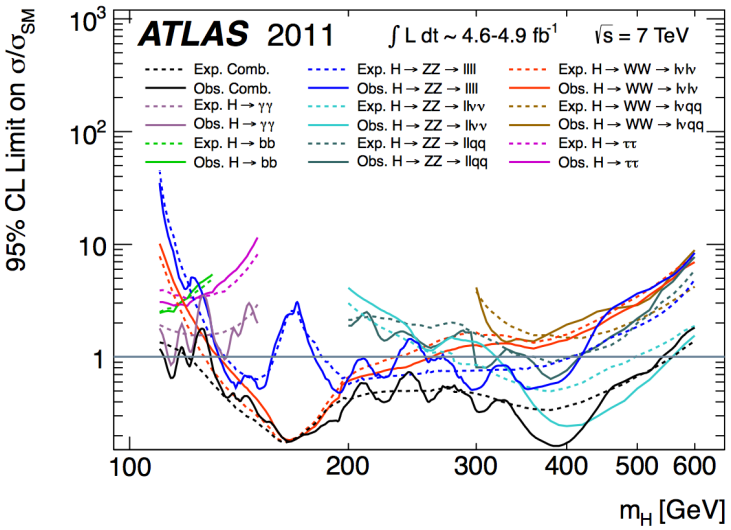


Figure 4.5 The observed (solid) and expected (dashed) 95% CL cross section upper limits for the individual search channels and the combination, normalized to the SM Higgs boson production cross section, as a function of the Higgs boson mass for the full Higgs boson mass hypotheses range.

The signal strength is called so that $\mu = 0$ corresponds to the background-only model and $\mu = 1$ is the standard model signal. It is convenient to separate the full list of parameters $\vec{\alpha}$ into the parameter of interest μ and the nuisance parameters $\vec{\theta}$: $\vec{\alpha} = (\mu, \vec{\theta})$.

For a given data set $\mathcal{D}_{\text{sim}}$ and values for the global observables $\mathcal{G}$ there is an associated likelihood function over μ and θ derived from combined model over all the channels including all the constraint terms[123].

$$L(\mu, \vec{\theta}; \mathcal{D}_{\text{sim}}, \mathcal{G}) = \mathbf{f}_{\text{tot}}(\mathcal{D}_{\text{sim}}, \mathcal{G}|\mu, \vec{\theta}) \ . \tag{4.1}$$

4.3.1 The test statistics and estimators of μ and $\vec{\theta}$

The maximum likelihood estimates (MLEs) $\hat{\mu}$ and $\hat{\vec{\theta}}$ and the values of the parameters that maximize the likelihood function $L(\mu, \vec{\theta})$ or, equivalently, minimize $-\ln L(\mu, \vec{\theta})$[123]. The dependence of the likelihood function on the data propagates to the values of the MLEs, so when needed the MLEs will be given subscripts to indicate the data set used.

For instance, $\hat{\vec{\theta}}_{\text{obs}}$ is the MLE of $\vec{\theta}$ derived from the observed data and global observables.

The conditional maximum likelihood estimate (CMLEs) $\hat{\hat{\vec{\theta}}}(\mu)$ is the value of $\vec{\theta}$ that maximizes the likelihood function with μ fixed; it can be seen as a multidimensional function of the single variable μ.

This procedure for choosing specific values of the nuisance parameters for a given value of μ, $\mathcal{D}_{\text{sim}}$, and $\mathcal{G}$ is often referred to as "profiling". Similarly, $\hat{\hat{\vec{\theta}}}(\mu)$ is often called "the profiled value of $\vec{\theta}$".

Given these definitions, we can construct the profile likelihood ratio

$$\lambda(\mu) = \frac{L(\mu, \hat{\hat{\vec{\theta}}}(\mu))}{L(\hat{\mu}, \hat{\vec{\theta}})} \ , \tag{4.2}$$

which depends explicitly on the parameter of interest μ, implicitly on the data $\mathcal{D}_{\text{sim}}$ and global observables $\mathcal{G}$, and is independent of the nuisance parameters $\vec{\theta}$ (which have been eliminated via "profiling").

In any physical theory the rate of signal events is non-negative, thus $\mu \geq 0$. However, it is often convenient to allow $\mu < 0$ (as long as the pdf $f_c(x_c|\mu, \vec{\theta}) \geq 0$ everywhere).

In particular, $\hat{\mu} < 0$ indicates a deficit of events signal-like with respect to the background only and the boundary at $\mu = 0$ complicates the asymptotic distributions. A trick that is equivalent to requiring $\mu \geq 0$ while avoiding the formal complications of a boundary, which is to allow $\mu < 0$ and impose the constraint in the test statistic itself.

In particular, one defines $\tilde{\lambda}(\mu)$

$$\tilde{\lambda}(\mu) = \begin{cases} \frac{L(\mu,\hat{\hat{\vec{\theta}}}(\mu))}{L(\hat{\mu},\hat{\vec{\theta}})} & \hat{\mu} \geq 0, \\ \frac{L(\mu,\hat{\hat{\vec{\theta}}}(\mu))}{L(0,\hat{\hat{\vec{\theta}}}(0))} & \hat{\mu} < 0 \end{cases} \tag{4.3}$$

This is not necessary when ensembles of pseudo-experiments are generated with "Toy" MC techniques, but since they are equivalent, usually is written $\tilde{\lambda}$ to emphasize the boundary at $\mu = 0$.

For discovery the test statistic $\tilde{q}_0$ is used to differentiate the background-only hypothesis $\mu = 0$ from the alternative hypothesis $\mu > 0$:

$$\tilde{q}_0 = \begin{cases} -2\ln\tilde{\lambda}(\mu) & \hat{\mu} > 0 \\ 0 & \hat{\mu} \leq 0 \end{cases} \tag{4.4}$$

Note that $\tilde{q}_0$ is test statistic for a one-sided alternative and, if it is considered the parameter of interest $\mu \geq 0$, then it is equivalent to the two-sided test (because there are no values of μ less than $\mu = 0$).

For limit setting the test statistic $\tilde{q}_\mu$ is used to differentiate the hypothesis of signal being produced at a rate μ from the alternative hypothesis of signal events being produced at a lesser rate $\mu' < \mu$:

$$\tilde{q}_\mu = \begin{cases} -2\ln\tilde{\lambda}(\mu) & \hat{\mu} \leq \mu \\ 0 & \hat{\mu} > \mu \end{cases} = \begin{cases} -2\ln\frac{L(\mu,\hat{\hat{\vec{\theta}}}(\mu))}{L(0,\hat{\hat{\theta}}(0))} & \hat{\mu} < 0\ , \\ -2\ln\frac{L(\mu,\hat{\hat{\vec{\theta}}}(\mu))}{L(\hat{\mu},\hat{\hat{\vec{\theta}}})} & 0 \leq \hat{\mu} \leq \mu\ , \\ 0 & \hat{\mu} > \mu\ . \end{cases} \tag{4.5}$$

Note that $\tilde{q}_\mu$ is a test statistic for a one-sided alternative; it is a test statistic for a one-sided upper limit.

The test statistic $\tilde{t}_\mu$ is used to differentiate signal being produced at a rate μ from the alternative hypothesis of signal events being produced at a lesser or greater rate $\mu' \neq \mu$.

$$\tilde{t}_\mu = -2\ln\tilde{\lambda}(\mu)\ . \tag{4.6}$$

$\tilde{t}_\mu$ is a test statistic for a two-sided alternative (as in the case of the Feldman-Cousins[124] technique, though this is more general as it incorporates nuisance parameters).

Note that if we consider the parameter of interest $\mu \geq 0$ and we the test at $\mu = 0$ then there is no "other side" and we have $\tilde{t}_{\mu=0} = \tilde{q}_0$.

Finally, if one relaxes the constraint $\mu \geq 0$ then the two-sided test statistic is written t_μ or, simply, $-2\ln\lambda(\mu)$.

4.3.2 The distribution of the test statistic and p-values

The test statistic should be interpreted as a single real-valued number that represents the outcome of the experiment. More formally,it is a mapping of the data to a single real-valued number: $\tilde{q}_\mu : \mathcal{D}_{\rm sim}, \mathcal{G} \to \mathbb{R}$.

For the observed data the test statistic has a given value, $\tilde{q}_{\mu,\rm obs}$. If one were to repeat the experiment many times the test statistic would take on different values, thus, conceptually, the test statistic has a distribution.

Since the number of expected events $\nu(\mu, \vec{\theta})$ and the distributions of the discriminating variables $f_c(x_c|\mu, \vec{\theta})$ explicitly depend on $\vec{\theta}$ the distribution of the test statistic will also depend on $\vec{\theta}$.

Denoting this distribution as

$$f(\tilde{q}_\mu|\mu, \vec{\theta}) \ , \tag{4.7}$$

and it obtained analogous expressions for each of the test statistics described above.

The p-value for a given observation under a particular hypothesis $(\mu, \vec{\theta})$ is the probability for an equally or more 'extreme' outcome than observed assuming that hypothesis

$$p_{\mu,\vec{\theta}} = \int_{\tilde{q}_{\mu,\rm obs}}^{\infty} f(\tilde{q}_\mu|\mu, \vec{\theta})\, d\tilde{q}_\mu \ . \tag{4.8}$$

The logic is that small p-values are evidence against the corresponding hypothesis. In Toy MC approaches, the integral above is really carried out in the space of the data $\int d\mathcal{D}_{\rm sim} d\mathcal{G}$.

The immediate difficulty arrives in the interest on μ but the p-values depend on both μ and $\vec{\theta}$. In the frequentist approach the hypothesis $\mu = \mu_0$ would not be rejected unless the p-value is sufficiently small *for all* values of $\vec{\theta}$. Equivalently, one can use the supremum p-value for over all $\vec{\theta}$ to base the decision to accept or reject the hypothesis at $\mu = \mu_0$.

$$p_\mu^{\rm sup} = \sup_{\vec{\theta}} \ p_{\mu,\vec{\theta}} \tag{4.9}$$

The key conceptual reason for choosing the test statistics based on the profile likelihood ratio is that asymptotically (i.e. when there are many events) the distribution of the profile likelihood ratio $\lambda(\mu = \mu_{\rm true})$ is independent of the values of the nuisance parameters. This follows from Wilks's theorem. In that limit $p_\mu^{\rm sup} = p_{\mu,\vec{\theta}}$ for all $\vec{\theta}$.

For results based on generating ensembles of pseudo-experiments using Toy MC techniques does not assume the form of the distribution $f(\tilde{q}_\mu|\mu, \vec{\theta})$, but knowing that it is approximately independent of $\vec{\theta}$ means that one does not need to calculate p-values for all $\vec{\theta}$ (which is not computationally feasible).

Since there may still be some residual dependence of the p-values on the choice of $\vec{\theta}$ we would like to know the specific value of $\vec{\theta}^{\rm sup}$ that produces the supremum p-value over $\vec{\theta}$. Since larger p-values indicate better agreement of the data with the

model, it is not surprising that choosing $\vec{\theta}^{\text{sup}} = \hat{\hat{\vec{\theta}}}(\mu)$ is a good estimate of $\vec{\theta}^{\text{sup}}$. This has been studied in detail by statisticians, and is called the Hybrid Re-sampling method and is referred to in physics as the 'profile construction'[125, 126, 127].

Based on the discussion above, the following p-value is used to quantify consistency with the hypothesis of a signal strength of μ:

$$p_\mu = \int_{\tilde{q}_{\mu,\text{obs}}}^{\infty} f(\tilde{q}_\mu|\mu, \hat{\hat{\vec{\theta}}}(\mu, \text{obs}))\, d\tilde{q}_\mu \ . \tag{4.10}$$

A standard 95% confidence-level, one-sided frequentist confidence interval (upper limit) is obtained by solving for $p'_{\mu_{up}} = 5\%$. For downward fluctuations the upper limit of the confidence interval can be arbitrarily small, though it will always include $\mu = 0$. This feature is considered undesirable since a physicist would not claim sensitivity to an arbitrarily small signal rate. The feature was the motivation for the modified frequentist method called CL_s[122].

To calculate the CL_s upper limit, we define p'_μ as a ratio of p-values:

$$p'_\mu = \frac{p_\mu}{1 - p_b} \ , \tag{4.11}$$

Where p_b is the p-value derived from the same test statistic under the background-only hypothesis

$$p_b = 1 - \int_{\tilde{q}_{\mu,obs}}^{\infty} f(\tilde{q}_\mu|0, \hat{\hat{\vec{\theta}}}(\mu = 0, \text{obs}))d\tilde{q}_\mu \ . \tag{4.12}$$

CL_s upper-limit on μ is denoted μ_{up} and obtained by solving for $p'_{\mu_{up}} = 5\%$. It is worth noting that while confidence intervals produced with the "CLs" method over cover, a value of μ is regarded as excluded at the 95% confidence level if $\mu < \mu_{up}$.

The amount of over coverage is not immediately obvious; however, for small values of μ the coverage is near 97.5% (due to $\langle p_b \rangle \approx \frac{1}{2}$) and for large values of μ the coverage is near the nominal 95% (due to $\langle p_b \rangle \approx 0$).

For the purposes discovery one is interested in compatibility of the data with the background-only hypothesis. Statistically, a discovery corresponds to rejecting the background-only hypothesis. This compatibility is based on the following p-value

$$p_0 = \int_{\tilde{q}_{0,obs}}^{\infty} f(\tilde{q}_0|0, \hat{\hat{\vec{\theta}}}(\mu = 0, \text{obs}))d\tilde{q}_0 \ . \tag{4.13}$$

This p-value is also based on the background-only hypothesis, but the test statistic $\tilde{q}_0$ is suited for testing the background-only while the test statistic $\tilde{q}_\mu$ in Equation 4.12 is suited for testing a hypothesis with signal.

It is customary to convert the background-only p-value into the quantile (or "sigma") of a unit Gaussian. This conversion is purely conventional and makes no assumption that the test statistic q_0 is Gaussian distributed.

The conversion is defined as:

$$Z = \Phi^{-1}(1 - p_0); \tag{4.14}$$

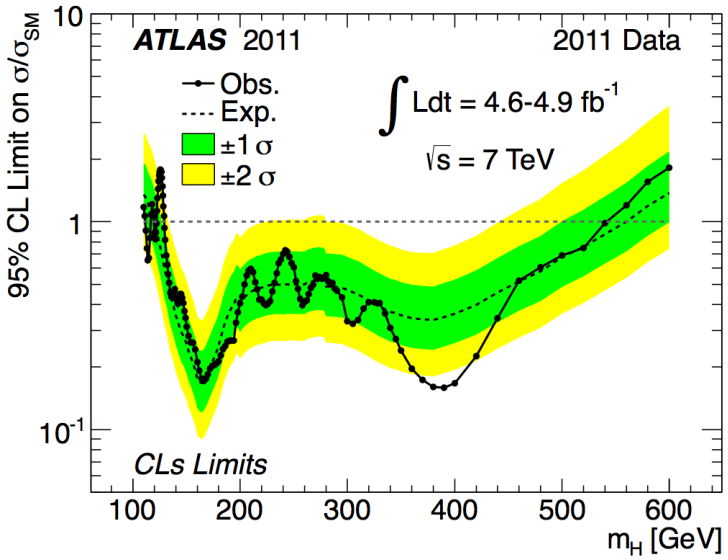


Figure 4.6 The observed (full line) and expected (dashed line) 95% CL combined upper limits on the SM Higgs boson production cross section divided by the SM expectation as a function of m_H in the full mass range considered. The dotted curves show the median expected limit in the absence of a signal and the green and yellow bands indicate the corresponding $\pm 1\sigma$ and $\pm 2\sigma$ intervals.

Where Φ^{-1} is the inverse of the cumulative distribution for a unit Gaussian. One says the significance of the result is $Z\sigma$ and the standard discovery convention is 5σ, corresponding to $p_0 = 2.87 \cdot 10^{-7}$.

See Figure 4.8 for the most uptodated ATLAS local probability p_0 for a background-only experiment to be more signal-like than the observation as a function of m_H for the combination of all channels[128]. The dashed curve shows the median expected local p_0 under the hypothesis of a Standard Model Higgs boson production at that mass (nuisance parameter values are taken from a fit to data with $\mu = 0$). The horizontal dashed lines indicate the $p-$values corresponding to significances of 0σ to 10σ.

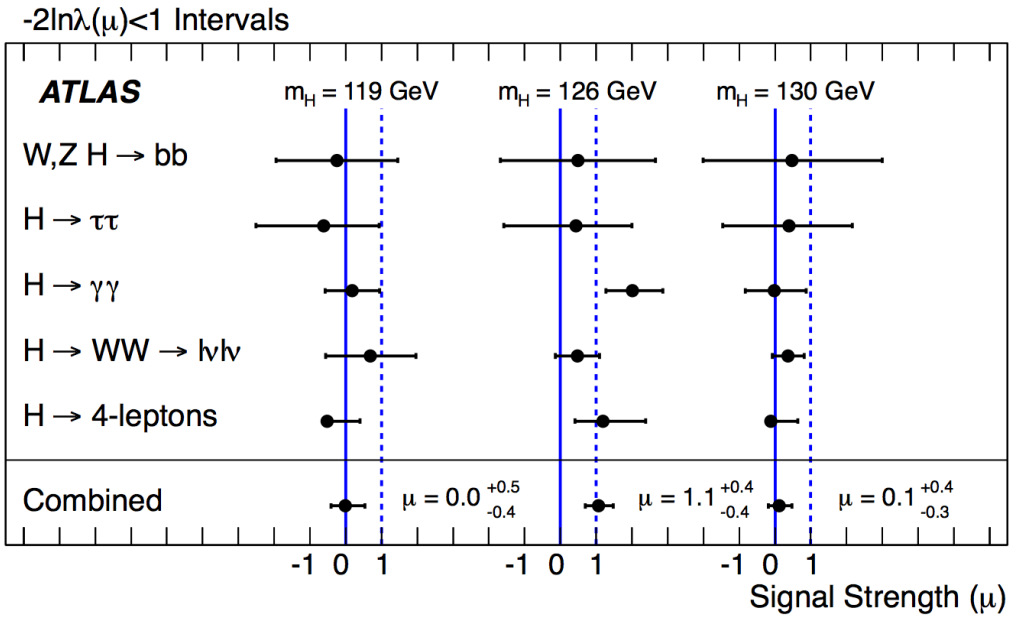


Figure 4.7 Summary of the individual and combined best-fit values of the strength parameter for three sample Higgs boson mass hypotheses 119 GeV, 126 GeV and 130 GeV.

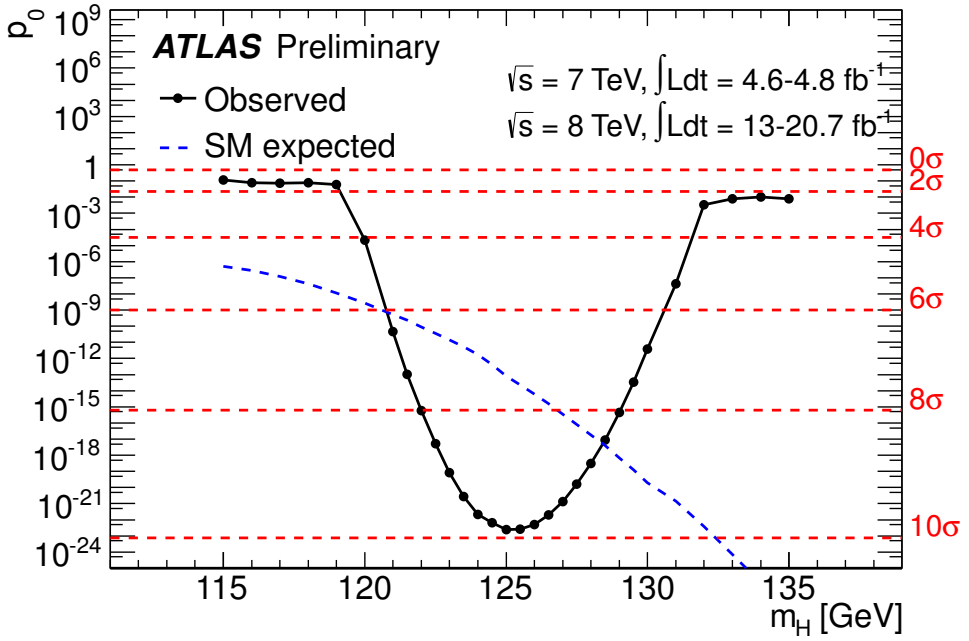


Figure 4.8 The local probability p_0 for a background-only experiment to be more signal-like than the observation as a function of m_H for the combination of all channels[128], updated March 2013.

Chapter 5

The $H \to ZZ^{(*)} \to l^+l^-q\bar{q}$ channel

In this chapter the analysis technique used for the study of the $H \to ZZ \to l^+l^-q\bar{q}$ channel (in the Low and High Mass ranges) will be described. Keeping in mind that for the moment this thesis have been written, the 2011 data was used for the so-called Low Mass $H \to ZZ^* \to \ell^+\ell^-q\bar{q}$ analysis and the 2012 data is currently in used into a complete (Low and High) $H \to ZZ \to l^+l^-q\bar{q}$ analysis, with its particular subdivisions. For that reason, next sections will have the description of *both analyses* with all the specifications relative to a particular year −2011 or 2012− introduced in a proper way.

Starting with the description of the MC and Data samples used in the finished 2011 analysis and the ongoing 2012 analyses (from here called: MC-2011, Data-2011, MC-2012 and Data-2012) following with the definitions in use of reconstructed objects and selection of the events, taking in consideration the final physics process to look at. A treatment of all the background sources and details on the systematic uncertainties affecting the measurements is given.

5.1 Data and Monte Carlo samples in 2011 and 2012

This section describes the data samples used in 2011 ($\sqrt{s} = 7$ TeV) and 2012 ($\sqrt{s} = 8$ TeV) analyses and the relevant MC generators used to model the signal and background processes. Some of the MC used have been modified varying from the 2011 to the 2012 analysis. This is due to the change of the LHC luminosity between the two mentioned years of data-taking.

5.1.1 Data sample

The 2011 data sample used, were recorded by the ATLAS detector during the LHC proton run, in which the protons have been collided at $\sqrt{s} = 7$ TeV. In the same way, the 2012 data was recorded at $\sqrt{s} = 8$ TeV. In both cases, the data are subsequently required to satisfy a number of conditions ensuring that ATLAS detector was fully operational with good efficiency while the data were collected (see Section 3.1). In 2011 the total integrated luminosity after these quality requirements is approximately $4.71 fb^{-1}$ with it's known with an accuracy of 3.9%.

m_H (GeV)	Gluon fusion σ (pb)	Vector-boson fusion σ (pb)	$\Gamma_{\ell\ell qq}/\Gamma_H$ (%)	Total $\sigma\cdot$ BR (fb)
120	16.65	1.279	0.222	39.8
125	15.32	1.222	0.370	61.2
130	14.16	1.168	0.559	85.6
135	13.11	1.117	0.789	112
140	12.18	1.069	0.965	128
145	11.33	1.023	1.11	137
150	10.58	0.9800	1.16	134
160	9.202	0.9043	0.583	59.8
170	7.786	0.8338	0.332	28.6
180	6.869	0.7684	0.847	64.5

Table 5.1 Cross sections in pb for the $H \to ZZ \to \ell\ell qq$ signal MC samples shown for the 2011 (7 TeV) analysis low mass range of Higgs boson. The cross sections are evaluated from theoretical calculations [106] for H production and Higgs boson branching fractions from [106].

In the case of 2012 data taking, the total integrated luminosity approximately $21.69 fb^{-1}$ (see Figure 2.1).

5.1.2 Signal samples

Simulated signal samples of $H \to ZZ^* \to \ell^+\ell^- q\bar{q}$and $H \to ZZ \to l^+l^-q\bar{q}$, where $\ell = e, \mu, \tau$ and $q = d, u, s, c, b$, have been generated for Higgs boson masses m_H = 120–180 GeV and m_H = 200–600 GeV respectively.

The POWHEG program is used to generate the simulated samples for both gluon fusion and vector boson fusion production mechanisms (See Section refLHCdiscovery).

Samples are available using POWHEG[101, 102, 103, 104] (interfaced to PYTHIA for hadronization), which includes matrix elements up to next-to-leading order.

Both gluon fusion and vector-boson fusion production mechanisms are taken into account by the matrix elements. In both MCs, PYTHIA is interfaced to PHOTOS[105] for initial state radiation. The normalisation of the signal MC samples is fixed using Higgs boson production cross sections and decay branching ratios taken from[138].

The cross sections for $H \to ZZ$ via gluon fusion are set to NNLO+NNLL+EW accuracy, and those for vector-boson fusion are evaluated at NLO+EW accuracy[64]. All signal samples correspond to luminosities that are much larger than the available luminosity in the data.

Table 5.1 contains the cross sections and the branching ratios for the signal samples for all the mass hypotheses considered in 2011 Low Mass analysis, and Table 5.2 contains the cross sections for the ongoing 2012 Low and High Mass analysis (also see Figures 4.1 and 4.2).

$\sqrt{s} = 7$ TeV		
m_H (GeV)	Gluon fusion σ (pb)	Vector-boson fusion σ (pb)
120	16.65	1.279
125	15.32	1.222
130	14.16	1.168
135	13.11	1.117
140	12.18	1.069
145	11.33	1.023
150	10.58	0.9800
155	9.886	0.9415
160	9.202	0.9043
165	8.378	0.8694
170	7.786	0.8338
175	7.299	0.7998
180	6.869	0.7684
$\sqrt{s} = 8$ TeV		
m_H (GeV)	Gluon fusion σ (pb)	Vector-boson fusion σ (pb)
120	21.13	1.649
125	19.52	1.578
130	18.07	1.511
135	16.79	1.448
140	15.63	1.389
145	14.59	1.333
150	13.65	1.280
155	12.79	1.231
160	11.95	1.185
170	10.12	1.098
180	8.874	1.015
200	7.127	0.8685
300	3.606	0.4408
400	2.924	0.2543
500	1.283	0.1561
600	0.5230	0.09688
700	0.2288	0.06330
800	0.1095	0.04365
900	0.05684	0.03164
1000	0.03163	0.02399

Table 5.2 Cross sections in *pb* for the $H \to ZZ \to \ell\ell qq$ signal MC samples shown for the Low-mass and High-mass ranges of Higgs boson masses at $\sqrt{s} = 7$ TeV and $\sqrt{s} = 8$ TeV. The cross sections are evaluated from theoretical calculations [106] for H production and Higgs boson branching fractions from [106].

5.1.3 Background samples

Several background processes give rise to final states with signatures similar to the above signal processes and are modelled with various event generators. Main background contributions come from Z + jets processes that are simulated using the ALPGEN MC generator[98].

For the 2011 analysis, relatively low mass same-flavour opposite-sign charged lepton pairs are selected, a significant background arises from $Drell - Yan$ l^+l^- hadro-production accompanied by multiple jets. The ALPGEN MC generator is also used for the simulation of the $Drell - Yan$ samples. Background samples for $t\bar{t}$, single top quark, and Wt production are simulated using the MC@NLO[101] event generator, which is interfaced to PYTHIA[35] for simulation of the underlying event. Diboson production is simulated using the HERWIG[36] event generator to account for off-shell bosons and cross-checked with MC@NLO. The cross sections of all simulated background samples are scaled ($\times$1.23 for Z + jets) to the highest order calculations available. The background from QCD multijet production in both the electron and muon channels is evaluated from the data and is discussed in Section 5.4.4. All generated events are fully simulated using the ATLAS detector simulation within the GEANT4 framework[97]. Additional pp interactions in the same and nearby bunch crossings (pile-up) are included in the simulation according to the corresponding data-taking recommendations, in this case MC samples are re-weighted to reproduce the observed distribution of the mean number of interactions per bunch crossing in the data.

Z+jets production

The $Z \to ee$, $Z \to \mu\mu$, and $Z \to \tau\tau$ background processes are simulated using ALPGEN program, which generates hard matrix elements for Z boson production with with additional numbers of partons p in the final state, where p runs from 0 to 5. The cross sections listed in table 5.3 include a k-factor of 1.23, in order to take into account NLO calculation[129, 130]. Ad hoc samples are used for the Z production in association with heavy-flavour jets, in which the b-quark production happens in the hard interaction and not only in the hadronization process. The possible overlapping events between the Z+ light jets and the Z+ heavy flavour samples are removed following the procedure described in [131, 142] which is based on looking at the opening angle between pairs of heavy-flavour quarks. In the Z+light flavour samples, events containing heavy-flavour pairs generated via parton showering are removed if $\Delta R > 0.4$; conversely, in the heavy-flavour sample, events are removed that contain pairs with $\Delta R < 0.4$ produced directly from the matrix element. The cross sections listed in table 5.3 (for the two LHC's center-of-mass energies under study), include an further scaling of 1.4 for the $Z + b$-jets samples in order to take into account NLO corrections[95][96]. The cross sections for the Z + jets samples used into the analysis are listed in Table 5.3.

Process	σ(fb) at $\sqrt{s} = 7$ TeV	σ(fb) at $\sqrt{s} = 8$ TeV
$Z + 0p, Z \to \ell\ell$	827 375	875 760
$Z + 1p, Z \to \ell\ell$	166 625	190 650
$Z + 2p, Z \to \ell\ell$	50 375	60 024
$Z + 3p, Z \to \ell\ell$	14 000	17 466
$Z + 4p, Z \to \ell\ell$	3 375	4 637
$Z + 5p, Z \to \ell\ell$	1 000	1 378
$Zb\bar{b} + 0p, Z \to \ell\ell$	8 208	9 128
$Zb\bar{b} + 1p, Z \to \ell\ell$	3 100	3 458
$Zb\bar{b} + 2p, Z \to \ell\ell$	1 113	1 232
$Zb\bar{b} + 3p, Z \to \ell\ell$	488	542

Table 5.3 Cross sections for the Z + jets samples generated using the ALPGEN MC generator, where p refers to the number of additional partons generated in the matrix element. The cross sections listed include a k-factor of 1.23. The cross sections for $Z \to ee$, $Z \to \mu\mu$, and $Z \to \tau\tau$ production are assumed to be the same. The cross sections for the $Z + b$-jets are further scaled up by a factor 1.4 following [95].

Drell-Yan + jets (low mass) production

In the Low Mass 2011 analysis, the production of low mass same-flavour opposite-sign charged lepton pairs (see section 5.3) of the event selected have a significant background contribution arises from Drell-Yan $\ell\ell$ hadro-production accompanied by multiple jets. The ALPGEN MC generator is also used for the simulation of the Drell-Yan + jets process, where the dilepton mass, in MC-2011, is required to be in the range $10 < m_{\ell\ell} < 40$ GeV, while for MC-2012, such a dilepton mass is required to be in the range $10 < m_{\ell\ell} < 60$ GeV; with the same setting as for the minimum parton transverse momentum, parton shower, matching algorithm, as for the regular Z+jet. Cross sections are tabulated in Table 5.4, including a 1.22 k-factor for MC-2011 and a 1.19 k-factor for MC-2012 to include NLO calculations. The Drell-Yan dilepton production in association with heavy flavour jets is simulated in a similar way as the $Z + b$-jets process: a dedicated ALPGEN Drell-Yan $+b$-jets sample is used to describe events in which the heavy flavour jets are generated from the matrix element, and the overlapping events are removed with the same procedure described above. The cross sections for the Drell-Yan+b-jets are further scaled up by a factor 1.4[142] for MC-2011 and MC-2012.

Process	σ(fb) at $\sqrt{s} = 7$ TeV	σ(fb) at $\sqrt{s} = 8$ TeV
$DY + 0p$	3 723 000	4 141 200
$DY + 1p$	107 000	130 900
$DY + 2p$	50 500	62 237
$DY + 3p$	10 200	13 447
$DY + 4p$	2 260	3 082
$DY + 5p$	561	825
$DYb\bar{b} + 0p$	20 260	22 690
$DYb\bar{b} + 1p$	3 160	3 855
$DYb\bar{b} + 2p$	1 180	1 451
$DYb\bar{b} + 3p$	566	821

Table 5.4 Cross sections for the Drell-Yan process +jets samples generated using the ALPGEN MC generator, where p refers to the number of additional partons generated in the matrix element. The 7 TeV cross sections listed include a k-factor of 1.22, while the 8 TeV cross sections listed include a k-factor of 1.19. The cross sections for ee, $\mu\mu$, and $\tau\tau$ production are assumed to be the same.

Top pair and single top production

Background samples of $t\bar{t}$, single top, and Wt production are simulated using the MC@NLO event generator [101] interfaced to JIMMY 4.31 [107] for simulation of the underlying event. The $t\bar{t}$ sample is filtered at generator level by requiring at least one lepton originating from a W boson with $p_{\mathrm{T}} > 1$ GeV. This ensures that only events with at least one leptonic (e, μ, τ) W boson decay are retained; excluding the case where both W bosons decay hadronically. Cross sections for the samples are given in Table 5.5 for the 2011 and 2012 MC samples.

Diboson production

The background coming from ZZ production is considered irreducible, however in a very small amount when the selection is applied for this analysis, due that arrived with the same final state as the signal process. Some contribution from WZ production is expected, when in the physics process, the Z boson decays into leptons and it's not possible to distinguish between the W and the Z bosons decaying hadronically because of the available resolution on jets. The background from WW process have been included, nevertheless its contribution is very small. Contribution from all the diboson final states (WW, WZ and ZZ) are taken into account and are modelled using the MC@NLO event generator [101] interfaced to JIMMY 4.31 [107] for simulation of the underlying event as done for the top samples. Cross sections for the diboson samples are given in Table 5.6 for the 2011 and 2012 MC samples.

$\sqrt{s} = 7$ TeV			
channel	σ (fb)	filter	σ_{filtered} (fb)
$t\bar{t}$	164 600	0.5562	91 551
single t (s-chan, $W \to e\nu$)	462	–	–
single t (s-chan, $W \to \mu\nu$)	455	–	–
single t (s-chan, $W \to \tau\nu$)	484	–	–
single t (t-chan, $W \to e\nu$)	7 117	–	–
single t (t-chan, $W \to \mu\nu$)	6 997	–	–
single t (t-chan, $W \to \tau\nu$)	7 448	–	–
Wt	14 600	–	–
$\sqrt{s} = 8$ TeV			
channel	σ (fb)	filter	σ_{filtered} (fb)
238.06*0.648 $t\bar{t}$	238 060	0.6480	154 263
single t (s-chan, $W \to e\nu$)	564	–	–
single t (s-chan, $W \to \mu\nu$)	564	–	–
single t (s-chan, $W \to \tau\nu$)	564	–	–
Wt	20 666	–	–

Table 5.5 Cross sections for the $t\bar{t}$ sample in the lepton-hadron (ℓh) or lepton-lepton ($\ell\ell$) decay mode and for the single top and Wt samples, all generated using the MC@NLO MC generator for MC-2011 and MC-2012 in current use. The cross sections are to NLO[108]. Cross sections are convoluted with branching fractions taken from the Particle Data Book [82]

Inclusive W boson production

Background samples for $W \to e\nu$, $W \to \mu\nu$, and $W \to \tau\nu$ produced in association with jets are simulated using the ALPGEN MC generator [99]. As for the other samples simulated using ALPGEN (Drell-Yan/Z with additional light and heavy flavour jets) separate samples are used to simulate hard matrix elements for W, $W + c$, and $W + b$ production with additional numbers of partons p in the final state, where p runs from 0 to 5. Again, to remove double counting between the inclusive and the specific c/b-jet samples, the overlap removal procedure based on the angular distribution of the heavy-flavour quarks, described in [100] is used. The cross sections, listed in Table 5.7, include a k-factor of 1.20 to make the inclusive W boson production cross section agree with NLO calculations [108].

Channel	σ(fb) at $\sqrt{s} = 7$ TeV	σ(fb) at $\sqrt{s} = 8$ TeV *(inclusive samples)*
$ZZ \to \ell\ell qq$	841.5	4 691.4
$ZZ \to \ell\ell\nu\nu$	160.4	
$ZZ \to \ell\ell\ell\ell$	27.0	
$ZZ \to \ell\ell\tau\tau$	27.0	
$ZZ \to \tau\tau\tau\tau$	6.8	
$ZZ \to \tau\tau\nu\nu$	80.3	
$WW \to \ell\nu\ell\nu$	2012	32 501.0
$WW \to \ell\nu\tau\nu$	2012	
$WW \to \tau\nu\tau\nu$	503.0	
$W^+Z \to \ell\nu qq$	1688.9	12 009.0
$W^+Z \to \ell\nu\ell\ell$	159.2	
$W^+Z \to qq\ell\ell$	489.4	
$W^+Z \to \tau\nu\ell\ell$	79.6	
$W^+Z \to \ell\nu\tau\tau$	79.6	
$W^+Z \to \tau\nu\tau\tau$	39.8	
$W^+Z \to qq\tau\tau$	249.2	
$W^-Z \to \ell\nu qq$	912.6	
$W^-Z \to \ell\nu\ell\ell$	86.1	
$W^-Z \to qq\ell\ell$	269.3	
$W^-Z \to \tau\nu\ell\ell$	43.0	
$W^-Z \to \ell\nu\tau\tau$	43.0	
$W^-Z \to \tau\nu\tau\tau$	21.5	
$W^-Z \to qq\tau\tau$	134.7	

Table 5.6 Cross sections for the ZZ samples (where $\ell = e, \mu, \tau$) generated using the MC@NLO MC generator. In the case on MC-2011, the cross section is evaluated in the range $66 < m_{\ell\ell} < 116$ GeV from theoretical calculations for ZZ production [108] convoluted with the Z boson branching fractions from [82].

QCD multijet production

The background from QCD multijet production in both the electron and muon channels is evaluated from the data and is discussed in 5.4.4. This is done since the available MC samples statistics may not be enough to describe the QCD activity in a proper way, which is expected to be a relevant background for the Low Mass Higgs analysis. For the High Mass analysis case, since this background have a small contribution in the final selection (see Figures 5.17 and 5.18) only in the initial steps of the analysis (as data/MC agreement) requires a correct inclusion of the QCD multijet process.

Process	σ(fb) at $\sqrt{s} = 7$ TeV	σ(fb) at $\sqrt{s} = 8$ TeV
$W + 0p, W \to e\nu$	8 300 000	8 044 000
$W + 1p, W \to e\nu$	1 560 000	1 579 200
$W + 2p, W \to e\nu$	453 000	477 200
$W + 3p, W \to e\nu$	122 000	133 930
$W + 4p, W \to e\nu$	30 900	35 622
$W + 5p, W \to e\nu$	8 380	10 553
$Wb\bar{b} + 0p$	56 800	55 682
$Wb\bar{b} + 1p$	42 900	45 243
$Wb\bar{b} + 2p$	20 800	23 246
$Wb\bar{b} + 3p$	7 960	11 144
$Wc\bar{c} + 0p$	153 000	150 190
$Wc\bar{c} + 1p$	126 000	132 680
$Wc\bar{c} + 2p$	62 500	71 807
$Wc\bar{c} + 3p$	20 400	30 264
$Wc + 0p$	518 000	807 890
$Wc + 1p$	192 000	267 610
$Wc + 2p$	51 000	69 823
$Wc + 3p$	11 900	20 547

Table 5.7 Cross sections for the W + jets samples generated using the ALPGEN MC generator, where p refers to the number of additional partons generated in the matrix element. The cross sections listed include a k-factor of 1.20. The cross sections for $W \to e\nu$, $W \to \mu\nu$, and $W \to \tau\nu$ production are assumed to be the same.

Pileup reweighting of Monte Carlo samples

Pileup affects this analysis in two main ways. First of all, extra jets from pileup events may be mistaken for signal jets. Moreover jets arising from the primary event and the missing E_T will gain extra energy from pileup events. In-time pileup refers to additional proton-proton interactions occurring in the same bunch crossing as the hard interaction of interest. Such interactions will produce extra soft particles and potentially reduce the efficiency for selecting signal events. During the 2011 LHC running, the number of interactions occurring per bunch crossing varied with time due to the changing machine parameters, including the beam intensity and the transverse size and number of bunches[134].

In addition to the in-time pileup there is also out-of-time pileup, which depends on the beam intensity in bunches preceding the one during which a recorded event occurs. This effect is accounted for in the MC samples, which assume a 50 ns bunch spacing, as was the case for the vast majority of the data used in this search. As the out-of-time pileup effects depend on the intensity of several prior bunches, the position of a bunch within the bunch train is also important. To model the effects of pileup, the MC samples for the above processes were simulated with a fixed distribution of additional minimum-bias interactions. This is subsequently reweighted to the distribution observed in the data using the pileup reweighting tool[135].

To verify the reweighting procedure, at least for the in−time pileup, the distribution of the number of primary vertices containing at least three tracks, N_{vtx}, is plotted after the selection of the leptonic Z boson candidate. Figure 5.1 shows the distribution of the number of primary vertices in data compared to the simulation after reweighting. It can be seen that the simulation provides a good description of the number of primary vertices over the majority of the range.

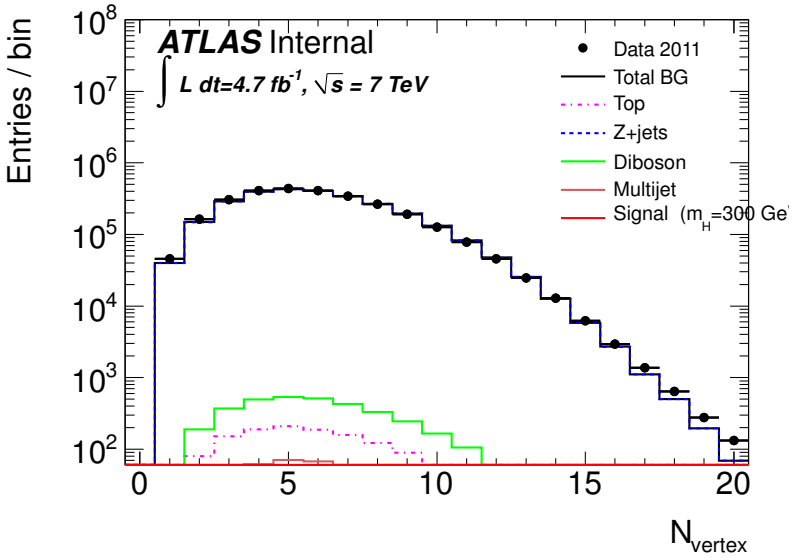


Figure 5.1 Comparison between data and MC simulation of the number of primary vertices, N_{vtx}, in events with two leptons within a mass window around m_Z, after the pileup reweighting.

5.2 Object selection in 2011 and 2012 data

Selection starts with requirements on the basic physics objects that will form the building blocks of the analysis. In this section the selection requirements for such a objects are explained with the calibration and smearing corrections to be applied to these objects as well as efficiency corrections applied to the MC that are directly related to the object selection are discussed[136].

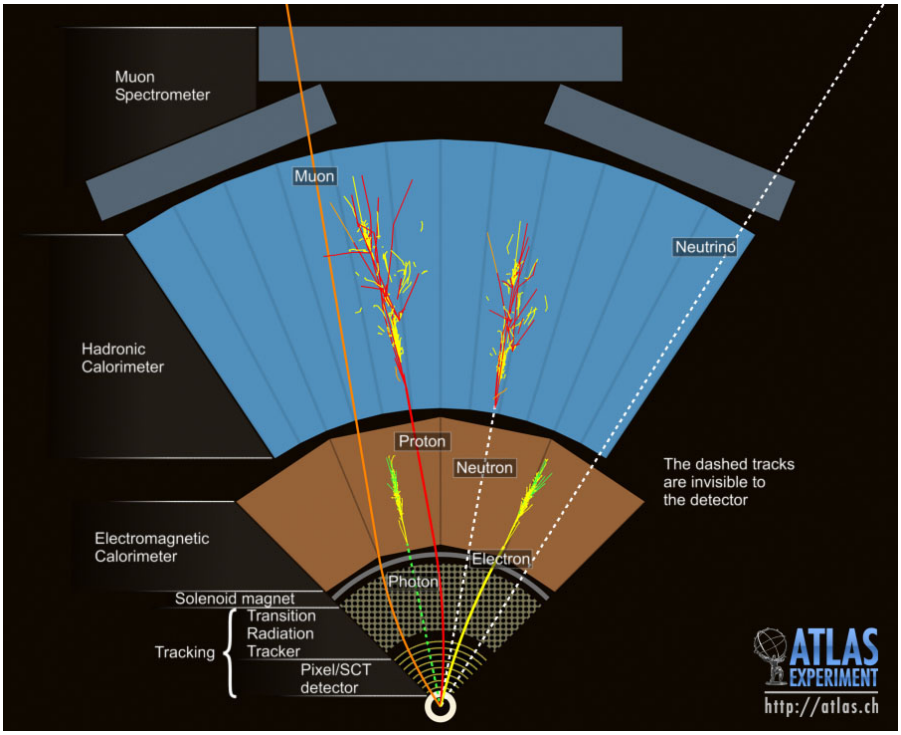


Figure 5.2 Representation of ATLAS detection of some objects to be used into the analysis.

5.2.1 Muons

In ATLAS four kinds of muon candidates are distinguished depending on the way they are reconstructed: stand-alone muons, combined muons, segment tagged muons, and calorimeter tagged muons (See Section 3.5.1 for details)[137].

Low Mass analysis at $\sqrt{s} = 7$ TeV. 2011	
Identification	Combined or segment-tagged STACO muons
Kinematic cuts	$p_{\mathrm{T}} > 7$ GeV $\lvert\eta\rvert < 2.5$
Inner Detector	$N_{\mathrm{hits}}^{\text{b-layer}} > 0$ (except where the muon passes an uninstrumented/dead area) $N_{\mathrm{hits}}^{\mathrm{pixel}} + N_{\mathrm{dead}}^{\mathrm{pixel}} > 1$ $N_{\mathrm{hits}}^{\mathrm{SCT}} + N_{\mathrm{dead}}^{\mathrm{SCT}} \geq 6$ $N_{\mathrm{holes}}^{\mathrm{pixel}} + N_{\mathrm{holes}}^{\mathrm{SCT}} < 3$ $\lvert\eta\rvert < 1.9$: $N_{\mathrm{tot}}^{\mathrm{TRT}} > 5$ and $N_{\mathrm{outliers}}^{\mathrm{TRT}} < 0.9 \times N_{\mathrm{tot}}^{\mathrm{TRT}}$ $\lvert\eta\rvert \geq 1.9$: If $N_{\mathrm{tot}}^{\mathrm{TRT}} > 5$, require $N_{\mathrm{outliers}}^{\mathrm{TRT}} < 0.9 \times N_{\mathrm{tot}}^{\mathrm{TRT}}$ where $N_{\mathrm{tot}}^{\mathrm{TRT}} = N_{\mathrm{hits}}^{\mathrm{TRT}} + N_{\mathrm{outliers}}^{\mathrm{TRT}}$
Cosmic rejection	$\lvert d_0\rvert < 1$ mm $\lvert z_0\rvert < 10$ mm
Track isolation	$\sum_{\text{tracks}} p_{\mathrm{T}}(\Delta R < 0.2)/p_{\mathrm{T}}^{\mu} < 0.1$
Additional QCD suppression	$\lvert d_0\rvert/\lvert\sigma_{d_0}\rvert < 3.5$

Table 5.8 Summary of muon selection in 2011 Low Mass Analysis. N_{hits} (N_{holes}) represents the number of hits (missing hits) in a particular subdetector of the inner tracker, while N_{dead} refers to the number of dead sensors crossed by the muon in a particular subdetector.

Muons are identified using STACO family, as they have been the standard for physics analyses in ATLAS, while the tracks are reconstructed in the muon spectrometer using the Muonboy algorithm[138]. These tracks are then extrapolated to the beam pipe and an attempt is made to find a matching inner detector track. If such a match is found, a combined muon is formed incorporating the information from both detectors. To recover muons that did not leave a full track in the muon spectrometer, remaining inner detector tracks (that are not used in the STACO combination) are extrapolated to the muon spectrometer and "tagged" as muons if they can be matched to a track segment in the first station. Such muons are known as segment-tagged muons. Both combined and segment-tagged muons are used in this analysis. for the 2011 analysis, all muon candidates are required to have $|\eta| < 2.5$ (to keep them within the acceptance of the muon systems and inner detector). In the case of the 2012 analysis, the η window had been extended with the use of the StandAlone[137] muons in the range of $2.5 < |\eta| < 2.7$. For the Low Mass analysis the p_T of the muon is request to be > 7 GeV, while for the High Mass analysis, $p_T > 10$ GeV.

Quality criteria for the muons object definition have been defined by ATLAS Muon Combined Performance (MCP) group[26], the inner detector track associated to the muon is required to pass a series of additional cuts based on the number of hits and holes (absence of hits) in the various layers of the inner detector. Such quality criteria varies between the years (2011-2012) and the corresponding updates have been applied. Muons from cosmic rays are suppressed by requiring the impact parameter with respect to the primary vertex satisfy $|d_0| < 1mm$ and $|z_0| < 10mm$[1]. To avoid muons associated with jets, such as those originating from semi-leptonic decays of b hadrons, the candidates are required to be isolated by demanding that the sum of the inner detector track momenta in a cone $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.2$ around the muon (ignoring the track associated to the muon itself) be less than 10% of the that of the muon.

To further reduce background contamination from QCD processes (mostly $b\bar{b}$-pair production) a cut on the impact parameter significance for both muons, $|d_0|/|\sigma_{d_0}| < 3.5$ (Figure 5.3) have been applied, which maintains a very high efficiency for signal but significantly reduce muons from semi-leptonic b decays. Such cut was proved to be useful in 2011 −Low Mass− analysis, in the case of 2012 such a cut is still under investigation for each Higgs mass regime.

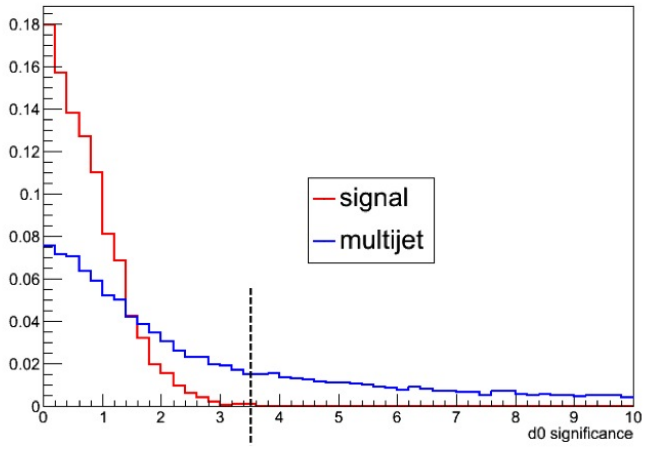


Figure 5.3 Distribution of the d_0 significance of the muons selected applying all the requirements described in Table 5.8 but the d_0 significance cut. The red line represents the signal sample for $m_H = 130$ GeV and the blue line represents the QCD distribution selected as described in 5.4.4. The two distribution are normalized to the same area.

Due to imperfection into the MC simulation (relevant in the case of the muon momentum resolution and the muon selection efficiency) in such MC, the muon momenta are smeared, and weights are applied to account for the difference in efficiency. The complete requirements for muons are shown in Table 5.8 for the 2011

[1]where d_0 and z_0 are the transverse and longitudinal impact parameters extrapolated at the primary vertex, respectively.

analysis and Tables 5.9 and 5.10 for the 2012 analysis, where in this last case the Low and High Mass selections are specified.

Low Mass analysis at $\sqrt{s} = 8$ TeV. 2012	
Identification	Combined or segment-tagged STACO muons Calorimeter tagged STACO muons Standalone muons STACO muons
Kinematic cuts (High Mass)	Staco CB+ST muons $p_T > 7$ GeV and $\|\eta\| < 2.5$ CaloTag muons $p_{\mathrm{T}} > 15$ GeV and $\|\eta\| < 0.1$ Staco StandAlone muons $p_T > 7$ GeV and $\|\eta\| > 2.5$, $\|\eta\| < 2.7$
Inner Detector	$N_{\mathrm{hits}}^{\mathrm{b\text{-}layer}} > 0$ (except where the muon passes an uninstrumented/dead area) $N_{\mathrm{hits}}^{\mathrm{pixel}} + N_{\mathrm{dead}}^{\mathrm{pixel}} > 0$ $N_{\mathrm{hits}}^{\mathrm{SCT}} + N_{\mathrm{dead}}^{\mathrm{SCT}} \geq 5$ $N_{\mathrm{holes}}^{\mathrm{pixel}} + N_{\mathrm{holes}}^{\mathrm{SCT}} < 3$ $\|\eta\| < 1.9$: $N_{\mathrm{tot}}^{\mathrm{TRT}} > 5$ and $N_{\mathrm{outliers}}^{\mathrm{TRT}} < 0.9 \times N_{\mathrm{tot}}^{\mathrm{TRT}}$ $\|\eta\| \geq 1.9$: If $N_{\mathrm{tot}}^{\mathrm{TRT}} > 5$, require $N_{\mathrm{outliers}}^{\mathrm{TRT}} < 0.9 \times N_{\mathrm{tot}}^{\mathrm{TRT}}$ where $N_{\mathrm{tot}}^{\mathrm{TRT}} = N_{\mathrm{hits}}^{\mathrm{TRT}} + N_{\mathrm{outliers}}^{\mathrm{TRT}}$
Cosmic rejection	$\|d_0\| < 1$ mm $\|z_0\| < 10$ mm
Track isolation	$\sum_{\mathrm{tracks}} p_{\mathrm{T}}(\Delta R < 0.2)/p_{\mathrm{T}}^{\mu} < 0.1$
Additional QCD suppression	$\|d_0\|/\|\sigma_{d_0}\| < 3.5$ (under study)

Table 5.9 Summary of muon selection in 2012 Low Mass Analysis. N_{hits} (N_{holes}) represents the number of hits (missing hits) in a particular subdetector of the inner tracker, while N_{dead} refers to the number of dead sensors crossed by the muon in a particular subdetector.

The p_{T} and η distributions of the muons with the selections described below are shown in Figures 5.4, 5.5 and 5.6. In general it can be seen that the simulation provides a reasonable description of the data distribution. An absolute normalization of the background components based on the integrated luminosity collected by ATLAS and the generation cross section of the MC samples is used for these plots for all the components except the QCD mutijets. The latter is estimated with data-driven methods described in Sec. 5.4.4.

High Mass analysis at $\sqrt{s} = 8$ TeV. 2012	
Identification	Combined or segment-tagged STACO muons Calorimeter tagged STACO muons Standalone muons STACO muons
Kinematic cuts (High Mass)	*Loose:* Staco CB+ST muons $p_T > 10$ GeV and $\|\eta\| < 2.5$ CaloTag muons $p_{\mathrm{T}} > 20$ GeV and $\|\eta\| < 0.1$ Staco StandAlone muons $p_T > 10$ GeV and $\|\eta\| > 2.5$, $\|\eta\| < 2.7$ *Medium:* Staco CB+ST muons $p_T > 25$ GeV and $\|\eta\| < 2.5$
Inner Detector	$N_{\mathrm{hits}}^{\mathrm{b\text{-}layer}} > 0$ (except where the muon passes an uninstrumented/dead area) $N_{\mathrm{hits}}^{\mathrm{pixel}} + N_{\mathrm{dead}}^{\mathrm{pixel}} > 0$ $N_{\mathrm{hits}}^{\mathrm{SCT}} + N_{\mathrm{dead}}^{\mathrm{SCT}} \geq 5$ $N_{\mathrm{holes}}^{\mathrm{pixel}} + N_{\mathrm{holes}}^{\mathrm{SCT}} < 3$ $\|\eta\| < 1.9$: $N_{\mathrm{tot}}^{\mathrm{TRT}} > 5$ and $N_{\mathrm{outliers}}^{\mathrm{TRT}} < 0.9 \times N_{\mathrm{tot}}^{\mathrm{TRT}}$ $\|\eta\| \geq 1.9$: If $N_{\mathrm{tot}}^{\mathrm{TRT}} > 5$, require $N_{\mathrm{outliers}}^{\mathrm{TRT}} < 0.9 \times N_{\mathrm{tot}}^{\mathrm{TRT}}$ where $N_{\mathrm{tot}}^{\mathrm{TRT}} = N_{\mathrm{hits}}^{\mathrm{TRT}} + N_{\mathrm{outliers}}^{\mathrm{TRT}}$
Cosmic rejection	$\|d_0\| < 1$ mm $\|z_0\| < 10$ mm
Track isolation	$\sum_{\mathrm{tracks}} p_{\mathrm{T}}(\Delta R < 0.2)/p_{\mathrm{T}}^{\mu} < 0.1$
Additional QCD suppression	$\|d_0\|/\|\sigma_{d_0}\| < 3.5$ (under study)

Table 5.10 Summary of muon selection in 2012 High Mass Analysis. N_{hits} (N_{holes}) represents the number of hits (missing hits) in a particular subdetector of the inner tracker, while N_{dead} refers to the number of dead sensors crossed by the muon in a particular subdetector.

5.2.2 Electrons

The electrons are identified starting from electromagnetic calorimeter clusters reconstructed with the standard ATLAS sliding window algorithm described in Section 3.4[138] that are matched to tracks in the inner detector. The electron candidates for the Low Mass analyses are required to satisfy the standard ATLAS Tight++ quality requirements (see Section 3.4.2). While for the High Mass selection the leading lepton has to be Medium++ and the second to be Loose++(see Table 5.13). To ensure a high reconstruction and trigger efficiency, the candidates are required to lie within the $|\eta_{cluster}| < 2.47$, i.e within the region of precision EM measurement, and have a (after energy correction/smearing) $p_T > 7$ GeV for the Low Mass analyses. For the High Mass selection such a p_T request is for the Medium++ electron $p_T > 25$ GeV and for the Loose++ $p_T > 10$ GeV.

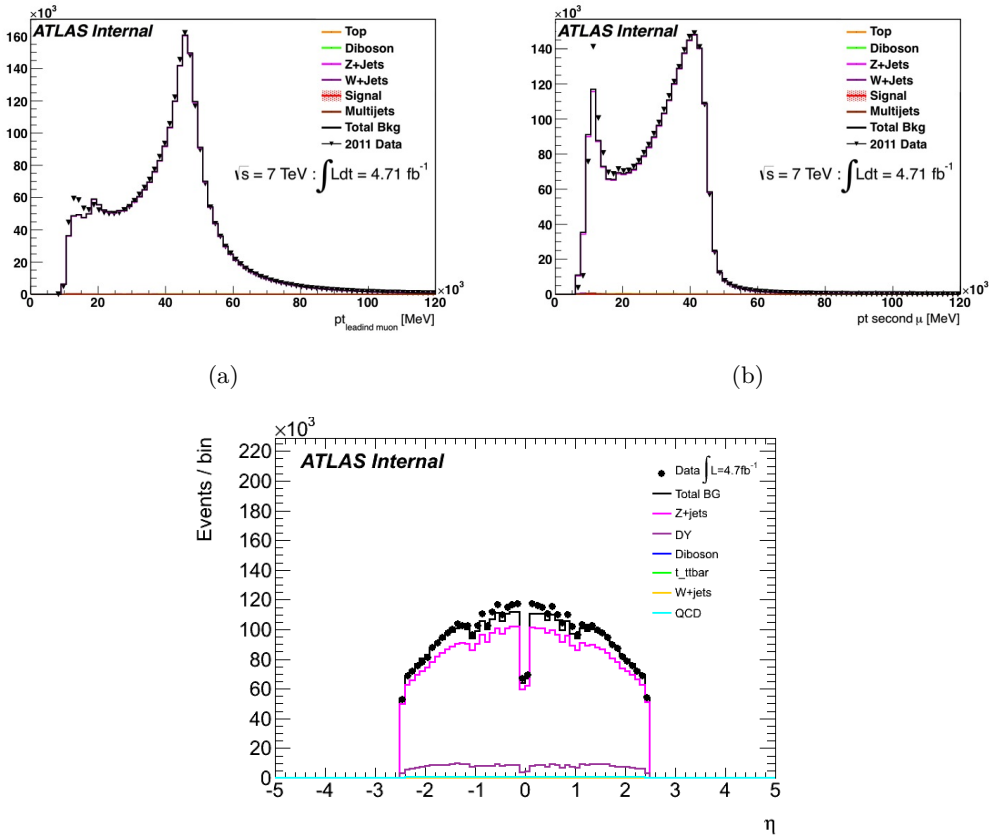
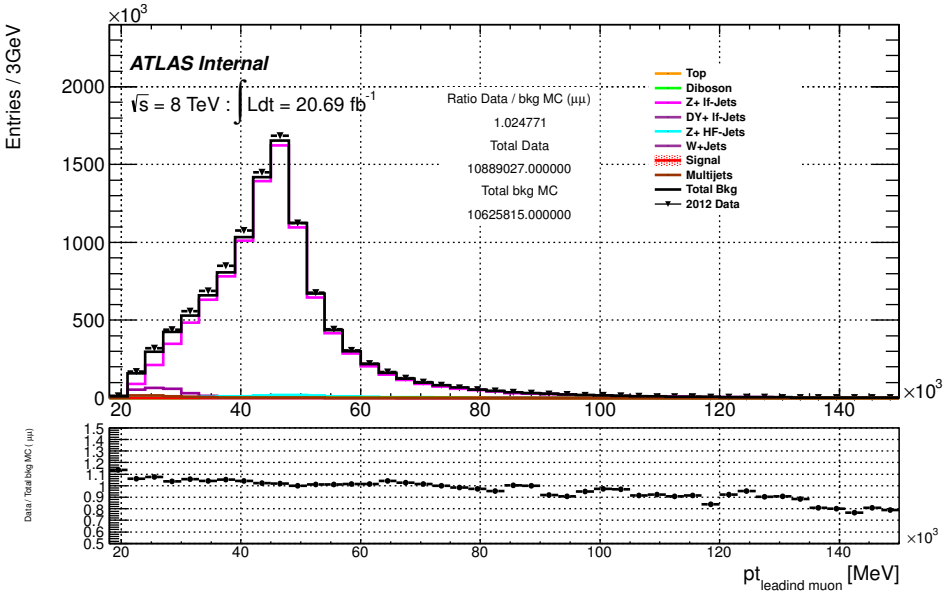
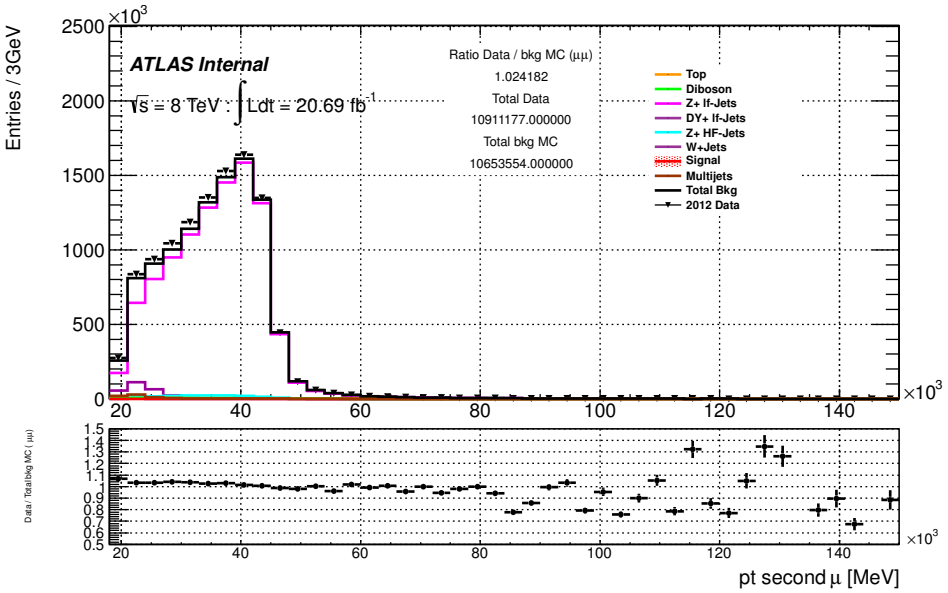


Figure 5.4 The p_{T} of the leading (a) and subleading muon (b) and η distribution of muon candidates (c) into the 2011 Low Mass Analysis with a total luminosity of $4.71 fb^{-1}$.

Isolation criteria have been applied: it is required that the sum of the inner detector track transverse momenta in a cone of $\Delta R < 0.2$ around the electron (excluding the track associated to the electron itself) is $< 10\%$ that of the electron itself. To avoid double counting and removing some fake electron from muon final state radiations electron candidates that lie within $\Delta R < 0.2$ of a selected muon (as defined above) are rejected. Electrons reconstruction involves both the ID and the calorimetric system, and hence the smearing and the other detector-related corrections should take into account also the measurement of the energy deposit in the calorimeter. In order to do this the standard ATLAS reconstruction includes cell-level energy corrections derived from data. However, residual corrections derived from the entire 2011 data set are available, to be applied at cluster level in order to have a proper energy measurement. These corrections are applied to the electron candidate energies in data and MC, and in addition to this, further smearing is applied to reproduce the electron energy resolution and the electron identification efficiency[27].



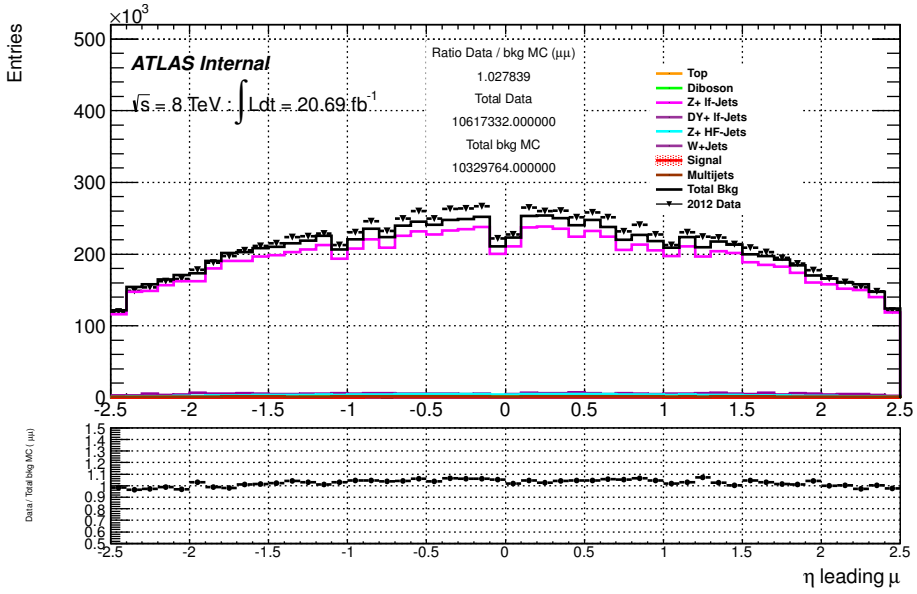
(a)



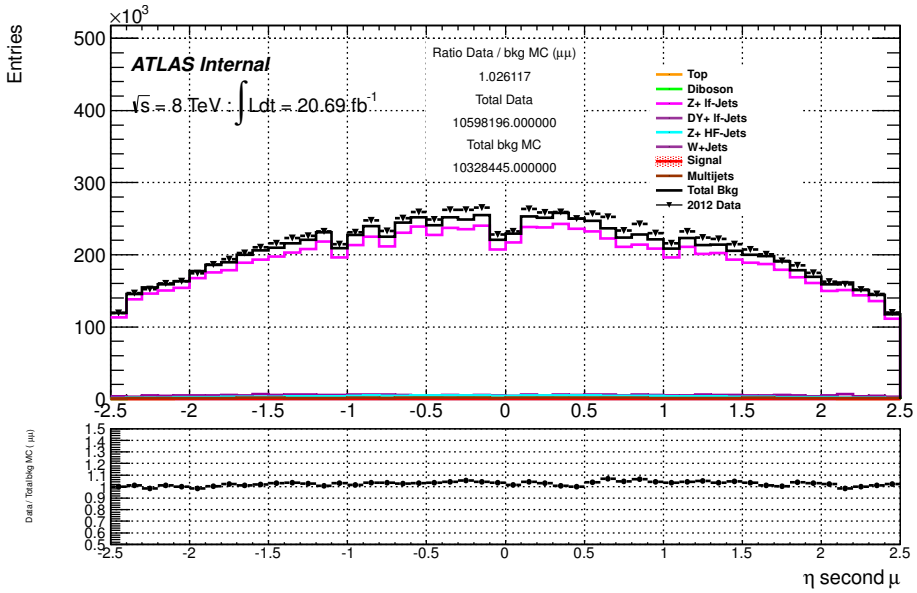
(b)

Figure 5.5 The p_{T} of the leading (a) and subleading muon (b) into the 2012 High Mass Analysis.

As for to the muon selection (see previous section) in the case of the 2011 Low Mass analysis to further reduce the background from QCD processes a cut on the (fakes, conversions and $b\bar{b}$-pair production) impact parameter significance, $|d_0|/|\sigma_{d_0}| < 6.5$ (Figure 5.10), have been applied in both electrons. This cut maintain a very high efficiency for signal but reduce electrons background from both semi-leptonic b decays and fakes. A similar optimization is under study for the 2012 analysis.



(a)



(b)

Figure 5.6 The η of the leading (a) and subleading muon (b) into the 2012 High Mass Analysis.

The electron selection is summarized in Table 5.11 for the 2011 analysis and Tables 5.12 and 5.13 for the 2012 analysis.

Low Mass analysis at $\sqrt{s} = 7$ TeV. 2011	
Identification	Author: Electron IsEM: Tight++
Kinematic cuts	$p_{\mathrm{T}} > 7$ GeV $\lvert\eta_{\mathrm{cluster}}\rvert < 2.47$
Track isolation	$\sum_{\mathrm{tracks}} p_{\mathrm{T}}(\Delta R < 0.2)/p_{\mathrm{T}}^{e} < 0.1$
Additional QCD suppression	$\lvert d_0\rvert/\lvert\sigma_{d_0}\rvert < 6.5$

Table 5.11 Summary of electron selection in the Low Mass 2011 analysis.

Low Mass analysis at $\sqrt{s} = 8$ TeV. 2012	
Identification	Author: GSF-Electron IsEM: Tight++ (under study)
Kinematic cuts	$p_{\mathrm{T}} > 7$ GeV $\lvert\eta_{\mathrm{cluster}}\rvert < 2.47$
Track isolation	$\sum_{\mathrm{tracks}} p_{\mathrm{T}}(\Delta R < 0.2)/p_{\mathrm{T}}^{e} < 0.1$
Additional QCD suppression	$\lvert d_0\rvert/\lvert\sigma_{d_0}\rvert < 6.5$ (under study)

Table 5.12 Summary of electron selection in the Low Mass 2012 analysis.

High Mass analysis at $\sqrt{s} = 8$ TeV. 2012	
Identification	Author: GSF-Electron IsEM: *Leading Electron:* Medium++ *Second Electron:* Loose++
Kinematic cuts	*Leading Electron:* $p_{\mathrm{T}} > 25$ GeV and $\lvert\eta_{\mathrm{cluster}}\rvert < 2.47$ *Second Electron:* $p_{\mathrm{T}} > 10$ GeV and $\lvert\eta_{\mathrm{cluster}}\rvert < 2.47$
Track isolation	$\sum_{\mathrm{tracks}} p_{\mathrm{T}}(\Delta R < 0.2)/p_{\mathrm{T}}^{e} < 0.1$
Additional QCD suppression	$\lvert d_0\rvert/\lvert\sigma_{d_0}\rvert < 6.5$ (under study)

Table 5.13 Summary of electron selection in the High Mass 2012 analysis.

In 2012, electron candidates must have a well-reconstructed ID track pointing to an electromagnetic calorimeter cluster and the cluster should satisfy a set of identification criteria that require the longitudinal and transverse shower profiles to be consistent with those expected for electromagnetic showers. Tracks associated with electromagnetic clusters are fitted using a Gaussian-Sum Filter[144] which allows for bremsstrahlung energy losses to be taken into account.

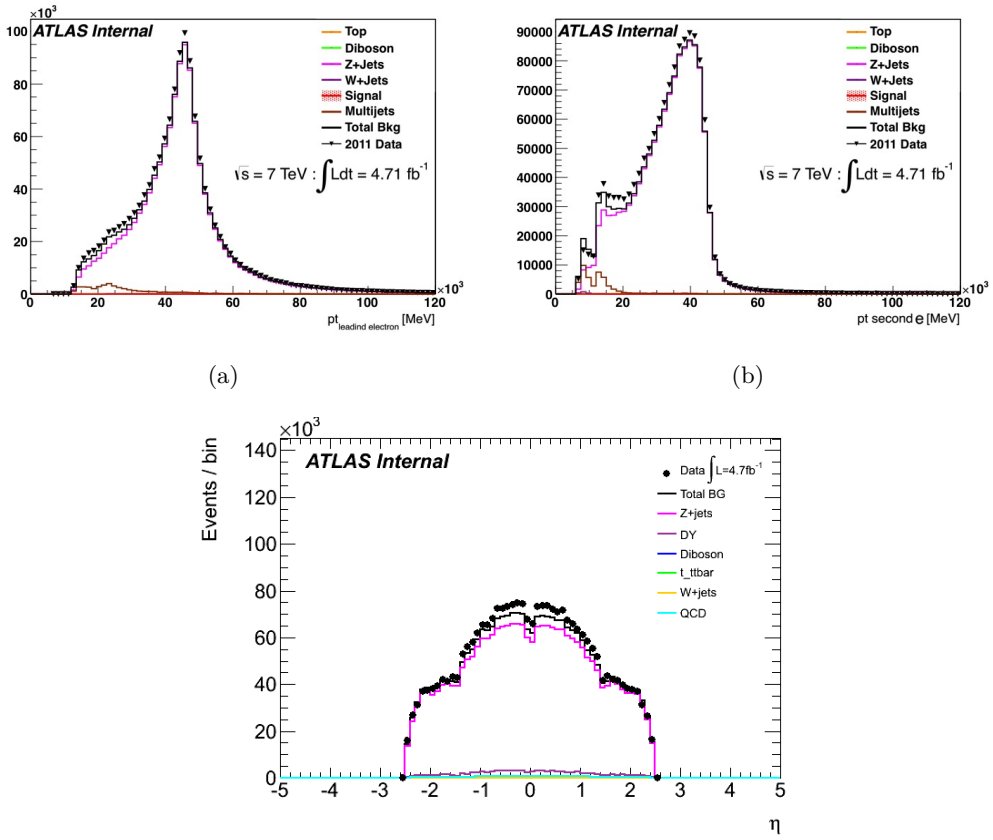
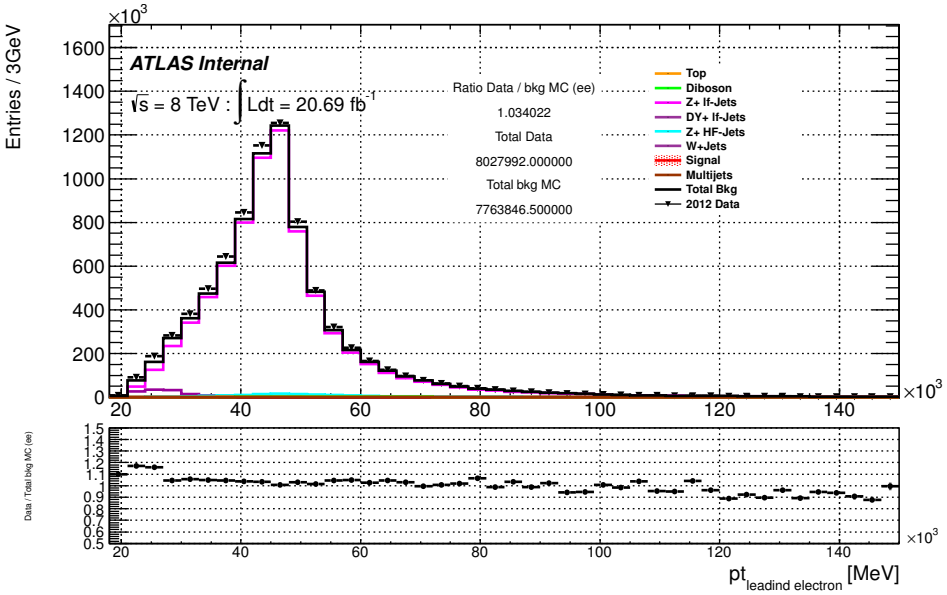


Figure 5.7 The p_{T} of the leading (a) and subleading electron (b) and η distribution of electron candidates (c) into the 2011 Low Mass Analysis.

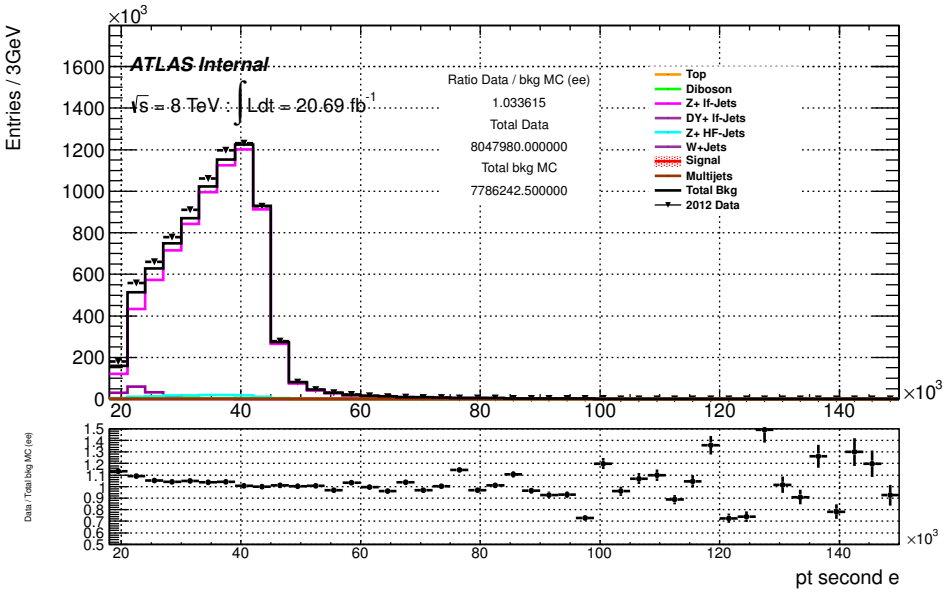
The electron identification in ATLAS is based on requirements on variables that provide good separation between isolated electrons and jets faking electrons (Section 3.4.2). Jets can deposit a large fraction of their energy in the hadronic calorimeter in contrast to electrons that deposit most of their energy in the electromagnetic calorimeter.

The comparison between the transverse energy in the hadronic and the electromagnetic calorimeter is a powerful discriminant against jets. Selections are also applied on the second sampling layer of the EM calorimeter where most of the electron energy is deposited (See Section 3.4).

The p_{T} and η distributions of the electrons selected according to the criteria explained above are shown in Figures 5.7, 5.8 and 5.9.



(a)

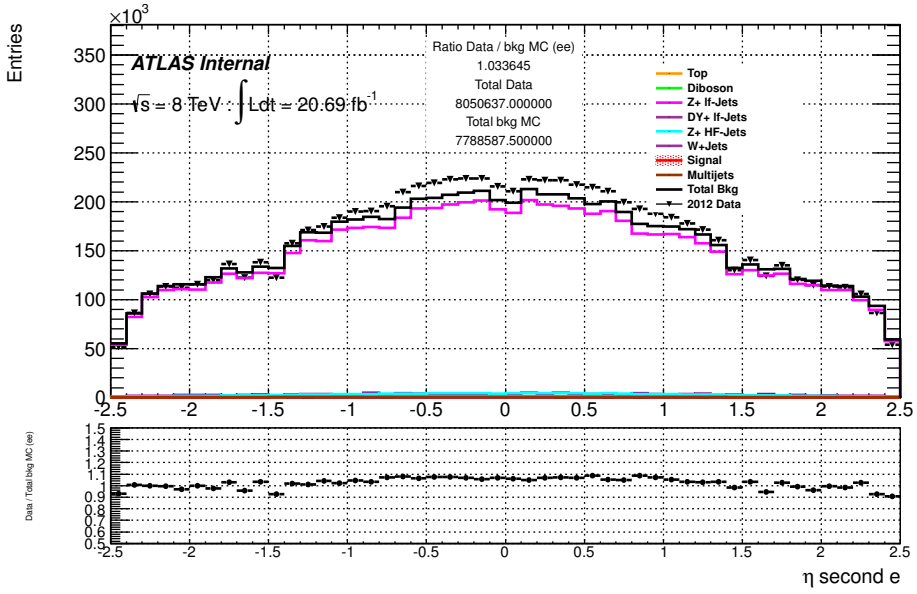


(b)

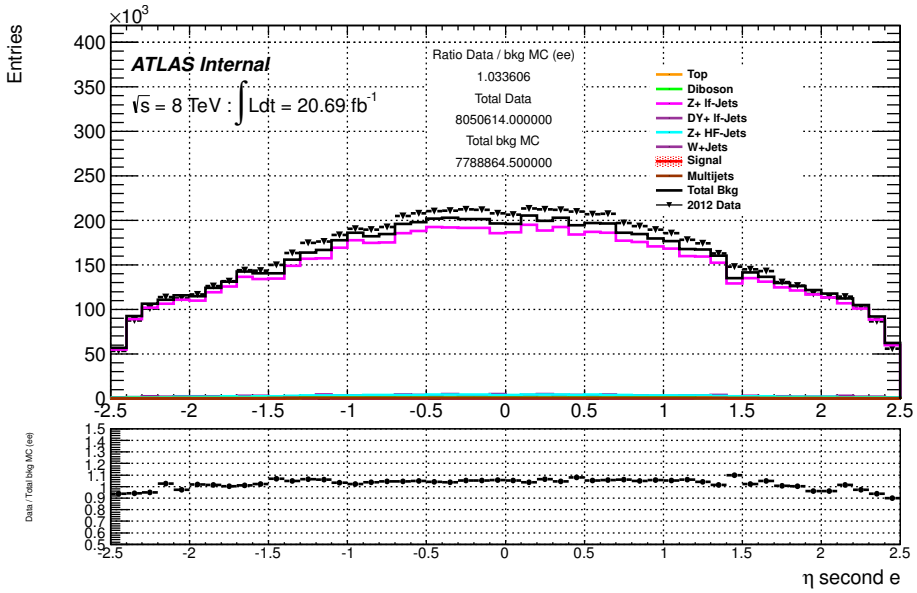
Figure 5.8 The p_{T} of the leading (a) and subleading electron (b) into the 2012 High Mass Analysis.

5.2.3 Jets

Jets are reconstructed from topological clusters [109] using an anti-k_T algorithm [110] with a distance parameter $\Delta R = 0.4$ (see Section 3.6). The jets are required to have $p_{\mathrm{T}} > 20$ GeV and are restricted to $|\eta| < 2.5$ in all the considered analyses, corresponding to the coverage of the ATLAS tracking detector. The jets are required to pass the "looser" quality cuts, which include requirements on the quality of the



(a)



(b)

Figure 5.9 The η of the leading (a) and subleading electron (b) into the 2012 High Mass Analysis.

calorimetric cluster.

Jets originating from pileup are removed by requiring that at least 75% of the tracks associated to the jet (within $\Delta R = 0.4$ around the jet axis) must originate from the primary vertex. This is implemented as a cut on the absolute value of the "jet vertex fraction"[2] $|\text{JVF}| > 0.75$ (See Figure 5.11).

[2]Variable that correspond to the fraction of track momentum associated to the jet under study.

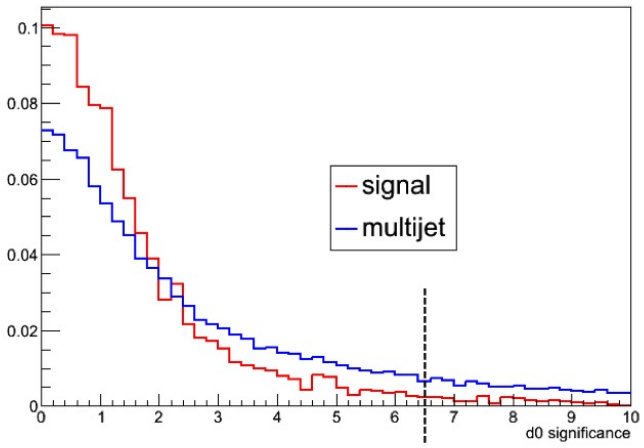


Figure 5.10 Distribution of the d_0 significance of the electrons selected applying all the requirements described in Table 5.11 but the d_0 significance cut. The red line represents the signal sample for $m_H = 130$ GeV and the blue line represents the QCD distribution selected as described in 5.4.4. The two distribution are normalized to the same area.

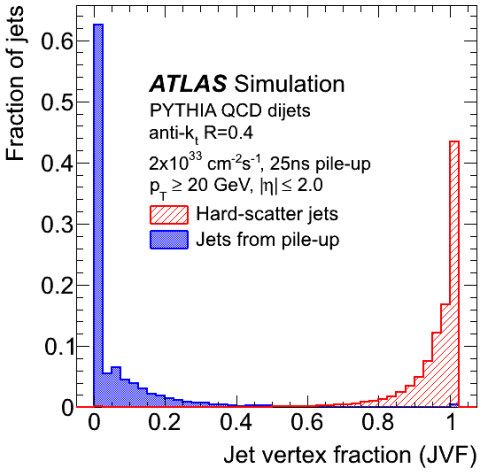
In order not to loose the information carried by jets with zero tracks (which may arise from hadronization of quarks coming from the signal process), also jets having JVF $= -1$ are included in the analysis[3][142].

The jet selection for the 2011 and 2012 analyses are summarized in Tables 5.14 and 5.15.

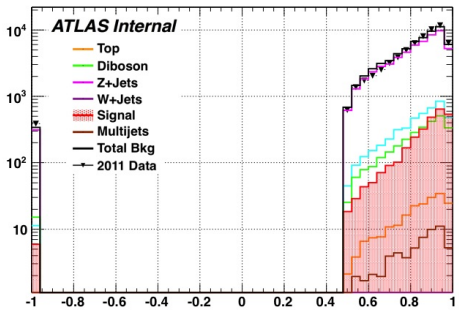
Identification	Anti-k_T $R = 0.4$ topological jets
Kinematic cuts	$p_{\mathrm{T}} > 20$ GeV $\lvert\eta\rvert < 2.5$
Quality	Looser quality cuts
Pileup	$\lvert\mathrm{JVF}\rvert > 0.75$

Table 5.14 Summary of jet selection in the Low Mass 2011 analysis.

[3]Jets with zero tracks may arise from real jets composed of solely neutral particles (which do not leave hits in the ID, and hence have no reconstructed track) or from jets falling in regions where the tracking efficiency is not optimal.



(a)



(b)

Figure 5.11 (a) Distribution of the JVF for jets produced in the hard scattering and those produced in the additional pileup interactions. (b) JVF distribution for the High Mass 2012 analysis. The signal sample is for $m_H = 400$ GeV.

Identification	Anti-k_T $R = 0.4$ topological jets
Kinematic cuts	*Veto:* $p_{\mathrm{T}} > 20$ GeV ($p_{\mathrm{T}} > 20$ GeV for $\|\eta\| > 2.5$) and $\|\eta\| < 4.5$ *Signal:* $p_{\mathrm{T}} > 20$ GeV and $\|\eta\| < 2.5$
Quality	Looser quality cuts
Pileup	$\|\mathrm{JVF}\| > 0.5$

Table 5.15 Summary of jet selection in the Low and High Mass 2012 analysis.

To avoid double-counting objects in the event, a jet is removed if an electron satisfying the criteria explained in section 5.2.2 is found within $\Delta R < 0.4$ around the jet axis.

The identification of b-jets

As was mentioned in Section 3.7, ATLAS $b-$tagging algorithms are used to distinguish jets containing decays of b hadrons from those containing only light quarks. A jet is then identified as a $b-$jet if the value of the $MV1$ discriminant is greater that a established operation point, giving a $b-$tagging efficiency of 70% (working point for 2011 and 2012 analyses) for $t\bar{t}$ events[92] while providing a light jet rejection of ≈ 140. Figure 5.12 shows the data-MC distibution of the MV1 efficiency as a function of the cut applied on its output value.

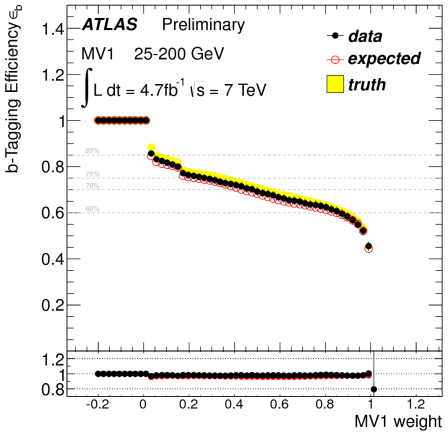


Figure 5.12 MV1 efficiency as a function of the cut applied on its output value.

5.2.4 Missing transverse energy

Missing transverse energy, E_T^{miss}, can be used to discriminate events with ν in the final state and hence improve the signal/background ratio. Since the $H \to ZZ^* \to \ell^+\ell^- q\bar{q}$ signal has small E_T^{miss}, an upper limit on the E_T^{miss} is applied to reduce the background from $t\bar{t}$, single t and Wt events, which are characterized by large $E_{\text{T}}^{\text{miss}}$ due to the presence of neutrinos generated in the top quark (or W) decay. As can be see from Figure 5.24 is dominated by top events at high values, and this feature is used to generate a top-enriched sample in order to check the MC modelization of this background, as explained in detail in section 5.4.2. The official recommendation in ATLAS for the reconstruction of $E_{\text{T}}^{\text{miss}}$ is to use the *RefFinal* algorithm illustrated in section 3.8. To obtain a systematic uncertainty on the measured E_T^{miss} value, the systematic scale and resolution variations applied to the muons are also applied in the E_T^{miss} measurement. For the electrons and jets, no explicit calibration is used in the E_T^{miss} determination described above; therefore, the difference between the systematically varied and/or smeared jet and electron momenta and the uncorrected momenta is used for the systematic uncertainty in E_T^{miss}[142].

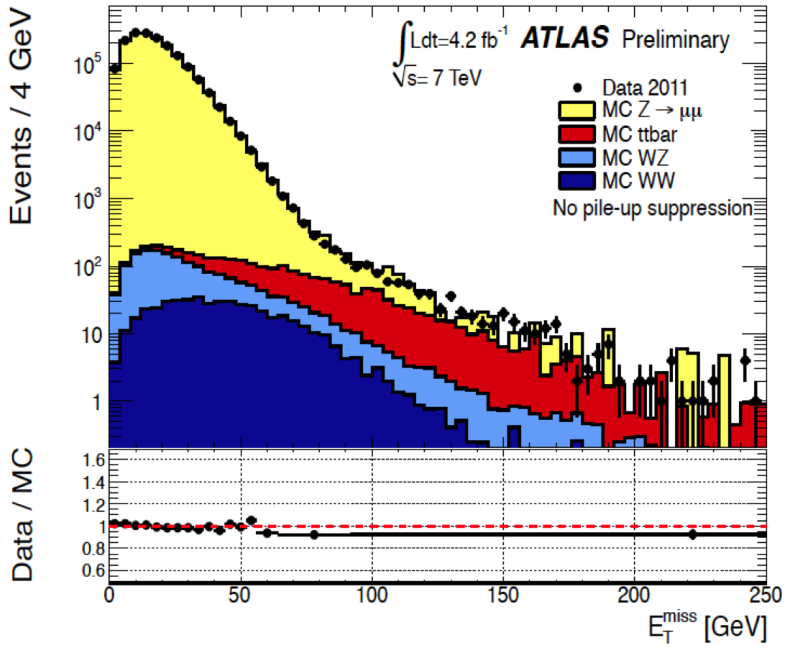
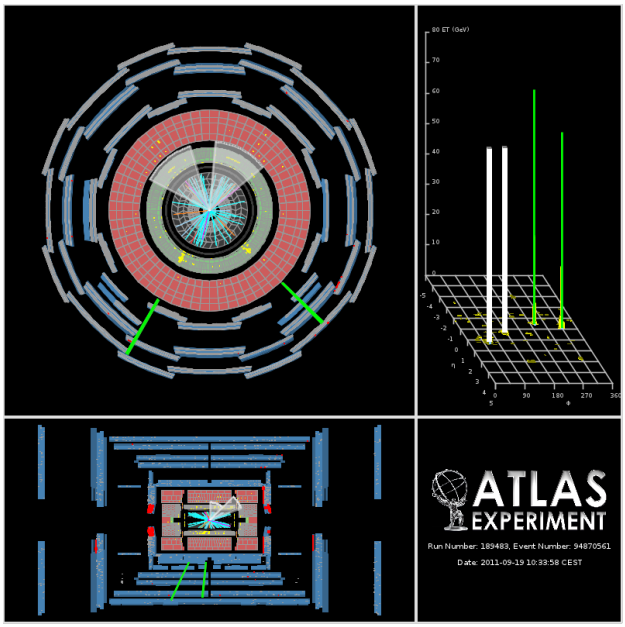


Figure 5.13 E_T^{miss} performance in $Z \to ll$ and $W \to e\nu$ events (systematic uncertainty from data/MC differences).

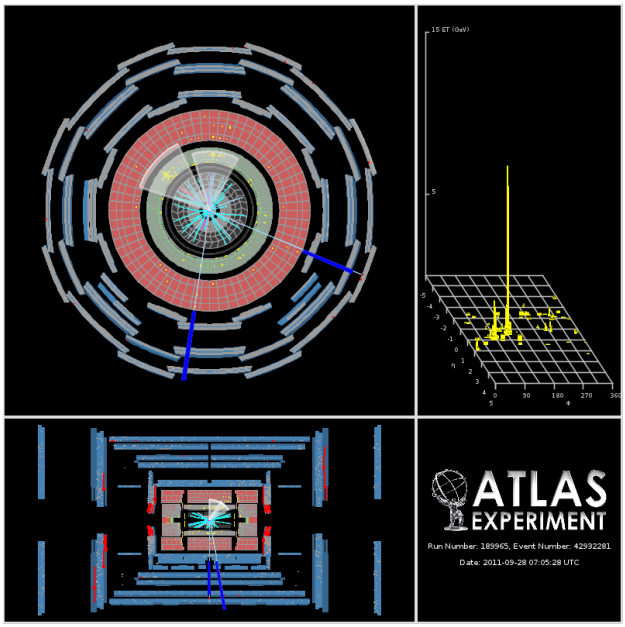
5.3 Event Selection in 2011 and 2012 data

The overall event selection strategy for the $H \to ZZ \to l^+l^-q\bar{q}$ channel is based on the request of a same flavour isolated dilepton (see Section 5.3.1) followed by the request of at least 2 good jets. When multiple jets in the final state arise the best jet pair has been selected using the procedure described in Section 5.3.3. The analysis have been further splitted in sub-channels depending on the number of $b-$jets in the final state ($0-1$ $b-$jet, 2 $b-$jets).

The full event selections for the 2011 and 2012 analyses are summarized in Tables 5.16, 5.17 and 5.18. At the moment that this dissertation have been written, the 2012 analyses are still under development.



(a)



(b)

Figure 5.14 (a) A candidate high-m_H $H \to ZZ \to eebb$ event with $m_{eebb} = 326$ GeV, and (b) a candidate high-m_H $H \to ZZ \to \mu\mu bb$ event with $m_{\mu\mu bb} = 332$ GeV. The two events contain two leptons (ee or $\mu\mu$) with $p_T > 20$ GeV and two identified $b-$jets with $p_T > 25$ GeV (grey cones). The left hand side shows the transverse (top) and longitudinal (bottom) views, while the top right plot shows the calorimeter energy versus η and ϕ.

Low Mass analysis at $\sqrt{s} = 7$ TeV. 2011	
Cut	Definition
Event quality	require that the detector is fully operational and that the reconstructed objects are of good quality
Trigger	lowest-p_{T} unprescaled triggers are required for both −single and double− lepton configurations: 20 GeV for single electron/muons and double 14 GeV for electron and 12 GeV muons.
Vertex	require the presence of at least one reconstructed vertex with 3 associated tracks
Lepton selection	exactly two same-flavour and opposite-sign leptons as defined in sections 5.2.1 and 5.2.2. Their p_{T} has to be consistent with the trigger thresholds (see inside text).
Missing transverse energy	$E_{\mathrm{T}}^{\mathrm{miss}} < 30$ GeV
Jet selection	at least two jets as defined in section 5.2.3 are required
Di-Jet Selection	Have been distinguished two event bins: - = 2 b-tags (b-tagged): Jets are identified as b-jets if $MV1 > 70\%$ operating point) - < 2 b-tags (untagged): select the two after the Kinematic Fitter answer, see 5.3.3 Events with more than two b-tags are vetoed.
Dilepton mass	$20 < m_{\ell\ell} < 76$ GeV
Dijet mass	$66 < m_{jj} < 115$ GeV

Table 5.16 Summary of the cuts applied to select the events used for the 2011 Low Mass analysis.

5.3.1 Dilepton event selection

The data used in this analysis have been passed a skimming procedure that requires at least two leptons, of any combination of electrons and muons[142]. The trigger signatures for the online selection of two-lepton events are single and dilepton triggers with the lowest unprescaled p_T threshold. All triggered events are required to contain a reconstructed primary vertex formed from at least three tracks ($p_T > 150$ MeV). To remove jets not originating from real in-time energy deposits, which arise from hardware problems, cosmic-ray showers, and LHC beam conditions, a jet cleaning cut is applied. In particular, these jets can give rise to fake missing transverse energy leading to indescribable tails in the E_T^{miss} distribution. To avoid this, any event containing a jet with $p_T > 20$ GeV which is identified as bad according to the looser cleaning criteria is rejected (this is done for both data and MC)[143].

Low Mass analysis at $\sqrt{s} = 8$ TeV. 2012	
Cut	Definition
Event quality	require that the detector is fully operational and that the reconstructed objects are of good quality
Trigger	lowest-p_{T} unprescaled triggers are required for both −single and double− lepton configurations: 20 GeV for single electron/muons and double 14 GeV for electron and 12 GeV muons.
Vertex	require the presence of at least one reconstructed vertex with 3 associated tracks
Lepton selection	exactly two same-flavour and opposite-sign leptons as defined in sections 5.2.1 and 5.2.2. Their p_{T} has to be consistent with the trigger thresholds (see inside text).
Missing transverse energy	$E_{\mathrm{T}}^{\mathrm{miss}} < 40$ GeV
Jet selection	at least two jets as defined in section 5.2.3 are required
Dilepton mass	$20 < m_{\ell\ell} < 76$ GeV
Di-Jet Selection	Have been distinguished three event bins: - 2 b-tags (b-tagged): Jets are identified as b-jets if $MV1 > 70\%$ operating point) - 1 b-tag: Jet is identified as b-jet if $MV1 > 70\%$ operating point) - 0 b-tags (untagged): select the two after the Kinematic Fitter answer, see 5.3.3 Events with more than two b-tags are vetoed.
Dijet mass	$66 < m_{jj} < 115$ GeV

Table 5.17 Summary of the cuts applied to select the events used for the 2012 Low Mass analysis.

Events are required to contain exactly two selected electrons or muons. Lepton transverse momenta are required to be consistent with the actual trigger firing the events. To ensure that the trigger item is approximately at its efficiency plateau, minimum lepton p_T is required to be greater than $12(14)GeV$ for muons (electrons) in case only dilepton trigger fired in the event. At least one lepton with p_T greater than 20 GeV is required if only the single lepton trigger fired.

High Mass analysis at $\sqrt{s} = 8$ TeV. 2012	
Cut	Definition
Event quality	require that the detector is fully operational and that the reconstructed objects are of good quality
Trigger	lowest-p_{T} unprescaled triggers are required for both −single and double− lepton configurations: 20 GeV for single electron/muons and double 14 GeV for electron and 12 GeV muons.
Vertex	require the presence of at least one reconstructed vertex with 3 associated tracks
Lepton selection	exactly two same-flavour leptons (opposite-sign for muons) as defined in sections 5.2.1 and 5.2.2. Their p_{T} has to be consistent with the trigger thresholds (see inside text).
Missing transverse energy	$E_{\mathrm{T}}^{\mathrm{miss}} < 60$ GeV
Jet selection	at least two jets as defined in section 5.2.3 are required
Dilepton mass	$83 < m_{\ell\ell} < 99$ GeV
Di-Jet Selection	Have been distinguished three event bins: - 2 b-tags (b-tagged): Jets are identified as b-jets if $MV1 > 70\%$ operating point) - 1 b-tag: Jet is identified as b-jet if $MV1 > 70\%$ operating point) - 0 b-tags (untagged): select the two after the Kinematic Fitter answer, see 5.3.3 Events with more than two b-tags are vetoed.
Dijet mass	$70 < m_{jj} < 105$ GeV
High mass cut ($m_H = 300$ GeV)	$p_T^{eachjet} > 45$ GeV $\Delta\phi_{ll} < \pi/2$, $\Delta\phi_{jj} < \pi/2$

Table 5.18 Summary of the cuts applied to select the events used for the 2012 High Mass analysis.

Additional QCD multijet reduction is obtained by requiring that $\Delta R > 0.3$ between any of the selected muons and any of the selected jets.
In 2011 and 2012 Low Mass analysis, the two leptons are required to be oppositely charged in both the muon and electron cases in order to reduce multijet background which produces both same sign and opposite sign pairs (for High Mass 2012 selection the opposite sign request is just applied to muons). The opposite sign requirement, in the electron channel, introduce an additional inefficiency ($\approx 1.5\%$) due to charge misidentification, but reduces significantly the QCD multi-jet background. Any event with additional selected leptons of either type is rejected to reduce background from WZ production. In the Low Mass analysis, one of the Z bosons is virtual (Z^*) and has been decided to select the Z^* decay leptonically because of the Z+jets background presented, and use as constrain the mass of the Z in the reconstruction of the jet pair. The invariant mass of the lepton pair must lie within the range $20 < m_{ll} < 76$ GeV for the Low Mass selection, and $83 < m_{ll} < 99$ GeV for the

High Mass. Some dilepton invariant mass distributions for the electron and muon channels are shown in Figure 5.16.

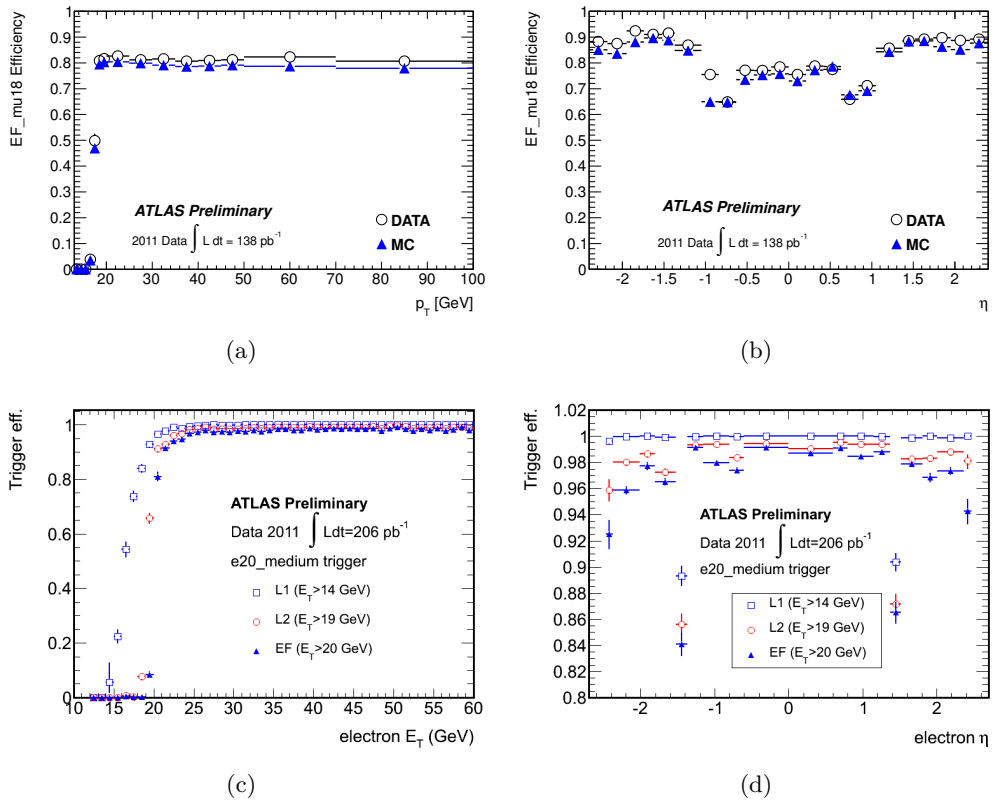
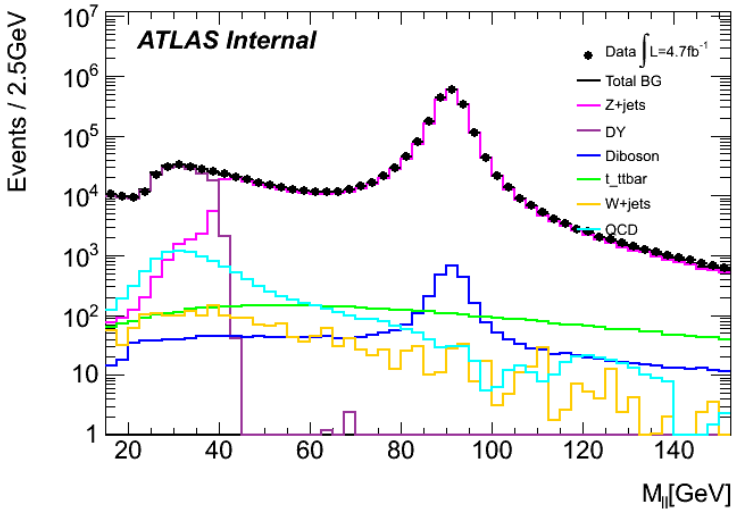
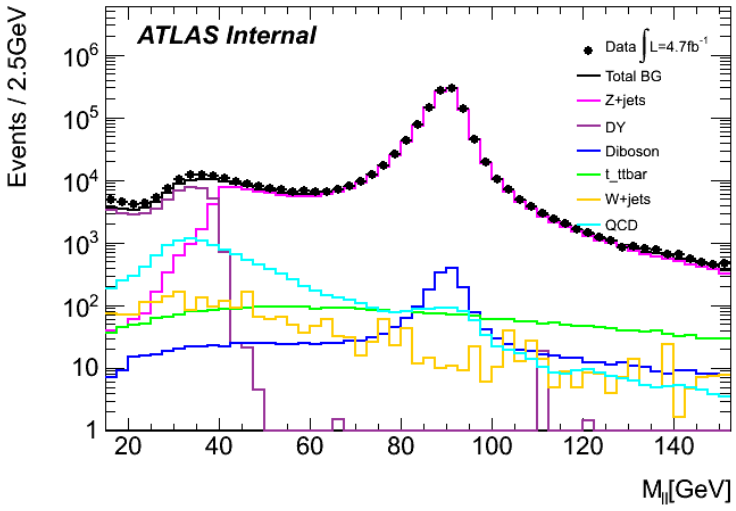


Figure 5.15 Efficiency of the single lepton trigger used for the analysis: the single muon trigger efficiency as a function of p_{T} (a) and η (b) and single electron trigger efficiency as a function of E_T (c) and η (d). For the electron trigger the efficiencies of the L1 and L2 triggers used to seed the final Event Filter chain are shown too.

Figure 5.15 shows some performances of the triggers used in this analysis.

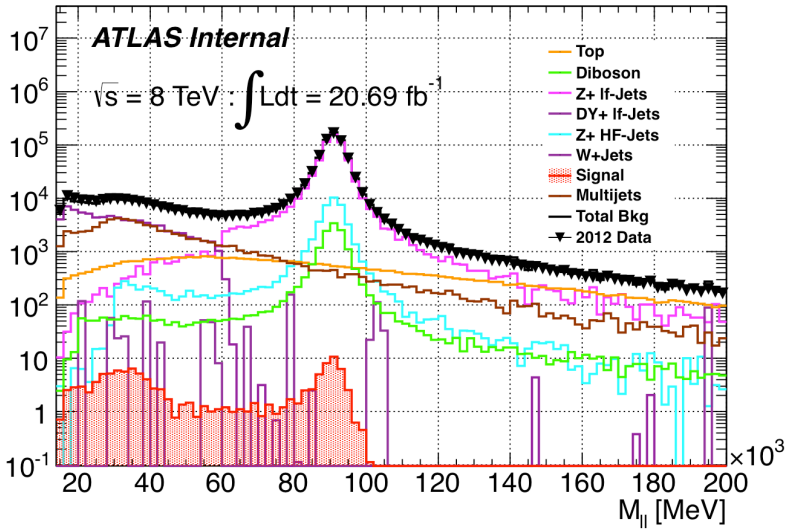


(a)

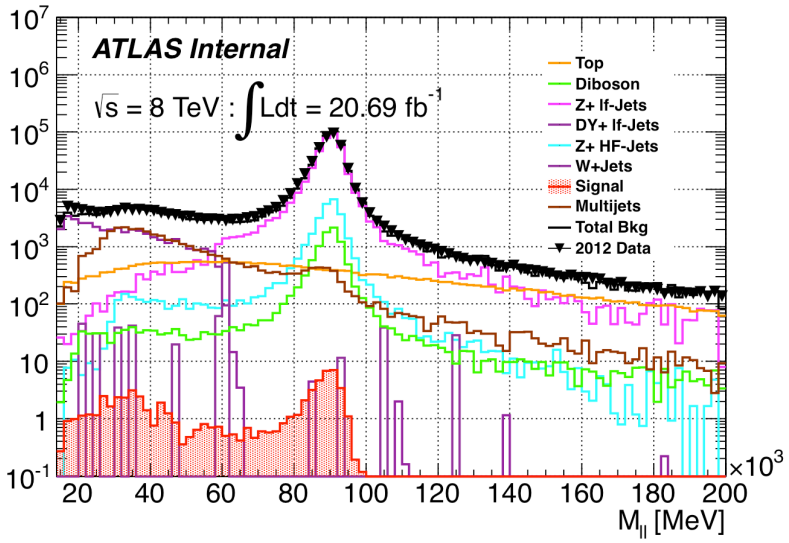


(b)

Figure 5.16 The dilepton invariant mass in the muon (a) and electron (b) channels for the Low Mass analysis unsing 2011 data.



(a)

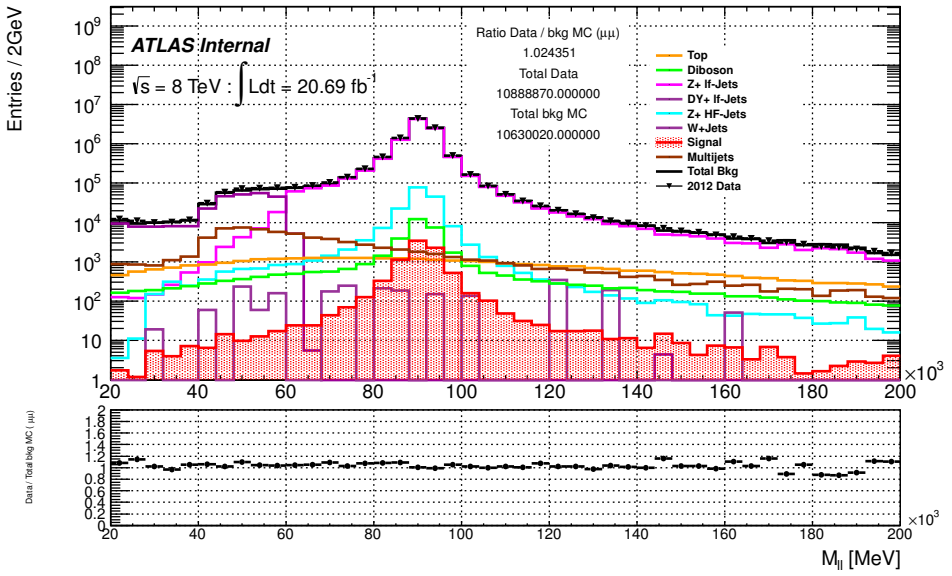


(b)

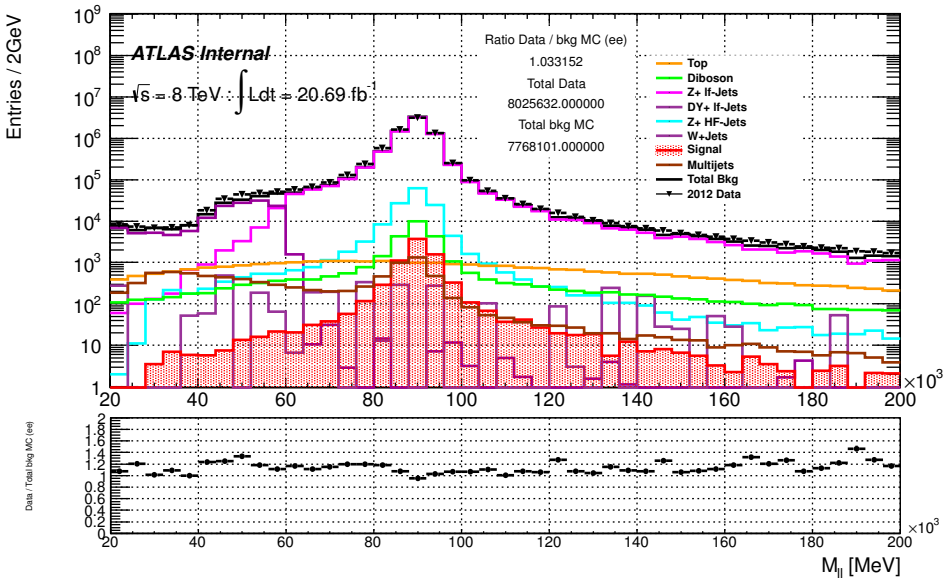
Figure 5.17 The dilepton invariant mass in the muon (a) and electron (b) channels for the Low Mass analysis unsing 2012 data.

5.3.2 $H \to ZZ^* \to \ell^+\ell^- q\bar{q}$ selection

After the dilepton request, the $H \to ZZ^* \to \ell^+\ell^- q\bar{q}$ candidates are further characterized by pair of jets resulting from $Z \to q\bar{q}$ decay, hence the presence of at least two jets as defined in section 5.2.3 is required. The jet multiplicities and jet p_{T} distribution after the dilepton selection are presented in Figure 5.20 for the 2011 Low Mass analysis (after the 20 GeV cut), and in Figures 5.21 and 5.22 for electron and muon channel respectively, for the 2012 High Mass analysis.



(a)



(b)

Figure 5.18 The dilepton invariant mass in the muon (a) and electron (b) channels for the High Mass analysis unsing 2012 data.

The lepton kinematic distributions after the dijet selection for the 2011 Low Mass analysis are displayed in Figure 5.23, showing a good consistency in both p_T and η distributions, thus giving confidence than a proper modelling of the Drell-Yan pairs kinematic can be obtained via the ALPGEN MC generator. A missing transverse energy requirement, $E_T^{miss} < 30$ GeV ($E_T^{miss} < 60$ GeV for 2012 High Mass selection), is then applied thus reducing mostly miss background from $t\bar{t}$ (Figure 5.24).

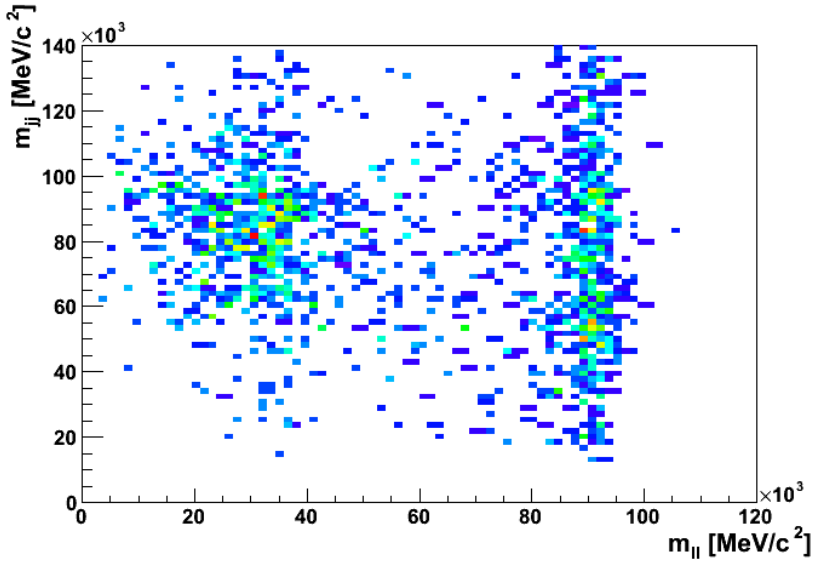


Figure 5.19 Invariant mass of the two leptons and of the two selected jets (see section 5.3.2 for details) in the event for $m_H = 130$ GeV: at least one of the two Z bosons is expected to be off mass-shell

Since about 22%[140, 141, 142] of our signal events contain $b-$jets, from $Z \to b\bar{b}$ decay, while $b-$jets are produced in Drell-Yan processes at a much lower rate ($\mathcal{O}(1\%)$)[142] the analyses have been splitted into "tagged" subchannel, containing events with exactly two $b-$tags, and "untagged" subchannel, containing events with zero or one $b-$tags, as is described in Tables 5.16, 5.17 and 5.18.

Figures 5.26 and 5.27 show a comparison of the observed distribution of the b-tag discriminant ($MV1$) to MC expectations for the leading (a) and subleading jet (b). Also showing the $b-$tagged jet multiplicities in the sample selected after the 2011 Low Mass E_T^{miss} requirement.

Dijet invariant mass are presented in Figure 5.28 5.28(a),5.28(b) for the 2011 Low Mass untagged case and the tagged case respectively. The untagged m_{jj} for the 2012 High Mass selection is shown too 5.28(c) and 5.28(d).

In the latter case an additional 5% correction on the jets p_T is applied in order to make the dijet mass for $Z \to b\bar{b}$ events peak (Figure 5.25) at the same value as for the light flavour dominated untagged case [113].

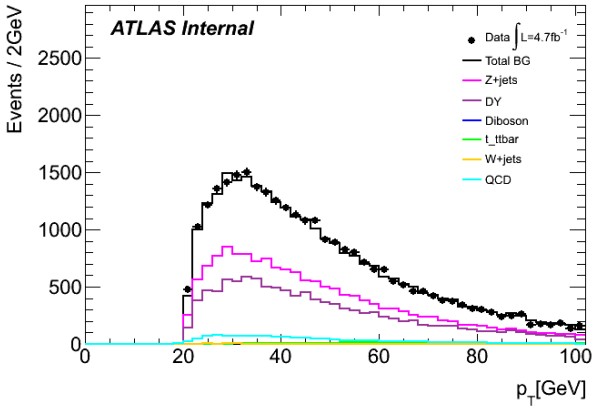
For $ZZ \to l^+l^-q\bar{q}$ Low Mass analysis (2011) the jet-jet is based in an algorithm that have been described in Section 5.3.3. A signal region (SR) and two side-band regions (SB1,SB2)[4] are defined using the invariant mass of the dijet system.

[4]We defined SB-region the union of the SB1 and SB2 side-bands.

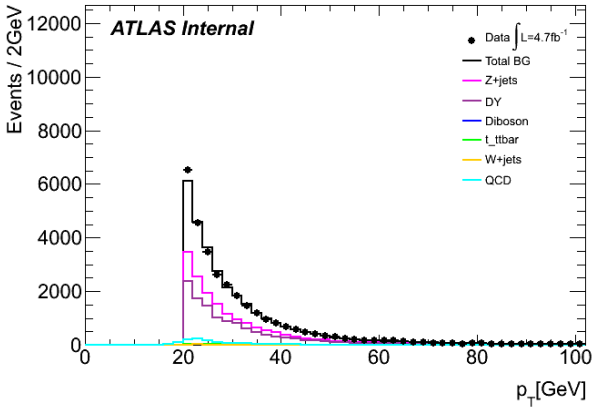
For the Low mass selections the signal region is defined for events where the dijet system has invariant mass in the range 60 GeV $< m_{jj} <$ 115 GeV. The two sidebands regions are statistically independent from the SR and correspond to the regions below and above the Z-peak region.

The "low mass" sideband, SB1, is defined with 40 GeV $< m_{jj} <$ 60 GeV while the "high mass" sideband, SB2, is defined with 115 GeV $< m_{jj} <$ 160 GeV. The two side-band region ranges are chosen such that they give similar event statistics in the two regions.

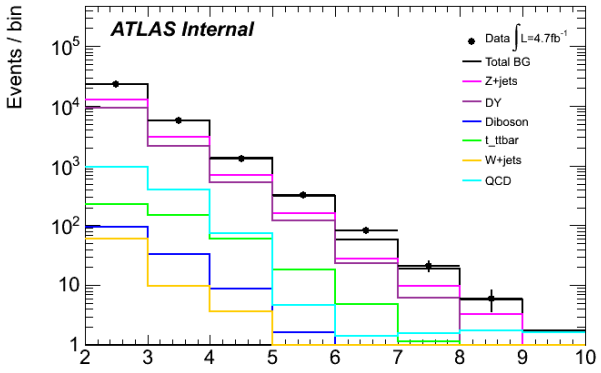
Summing the two channels (muon and electron), the resulting total analysis selection efficiency for the $H \to ZZ \to \ell\ell qq$ signal increases between 0.7–3% for the untagged selection and 0.05–0.25% for the tagged (with rest to each single channel) up to a maximum at roughly 170 GeV Higgs mass, dropping somewhat at the high end due to the requirement of an off-shell Z boson decaying to lepton pair. These efficiencies include $Z \to \tau^+\tau^-$ decays. For the individual channels, taking only into account muon and electron decays separately the efficiencies are shown in Figure 5.29(a) 5.29(b), respectively.



(a)

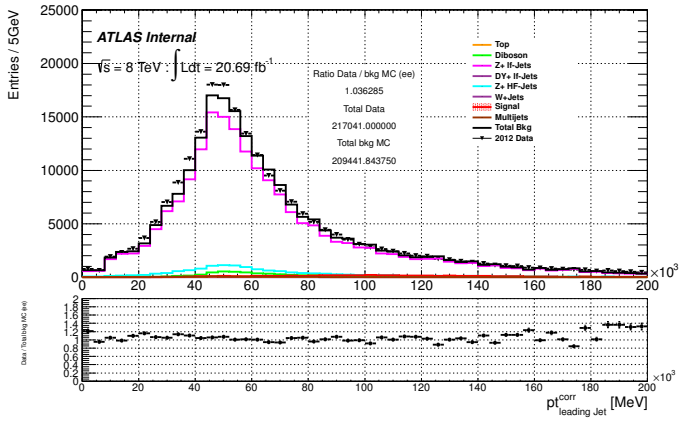


(b)

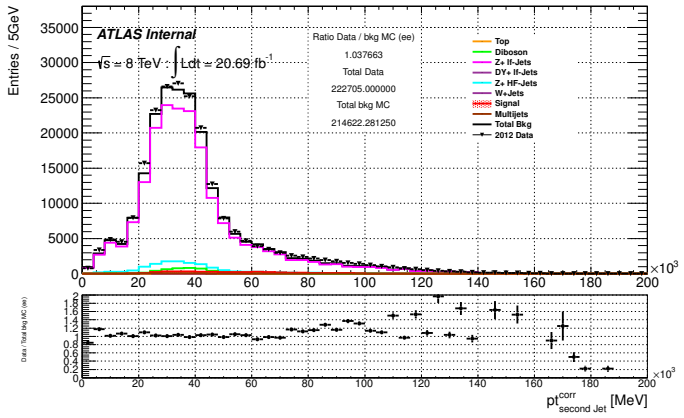


(c)

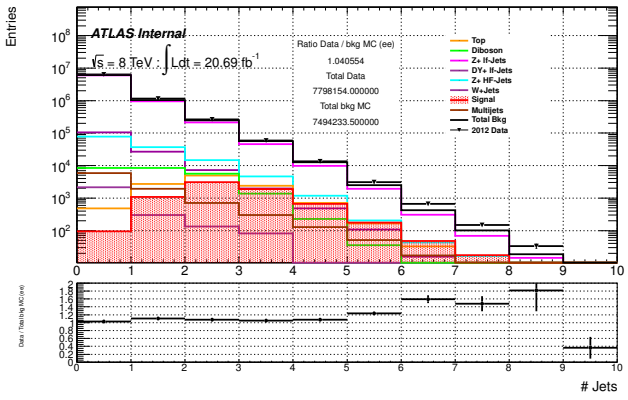
Figure 5.20 Distribution of the leading jet p_{T} (a), subleading jet p_{T} (b), and jet multiplicity after dilepton selection (c) for the combined electron and muon samples.



(a)

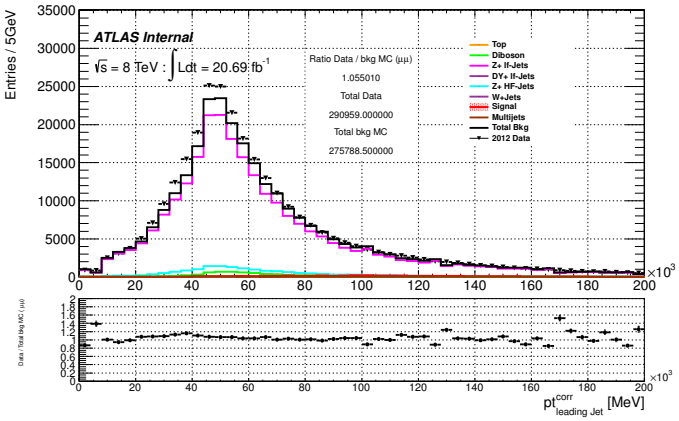


(b)

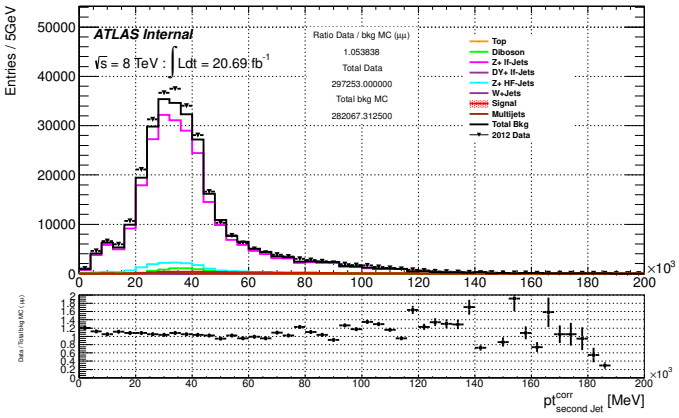


(c)

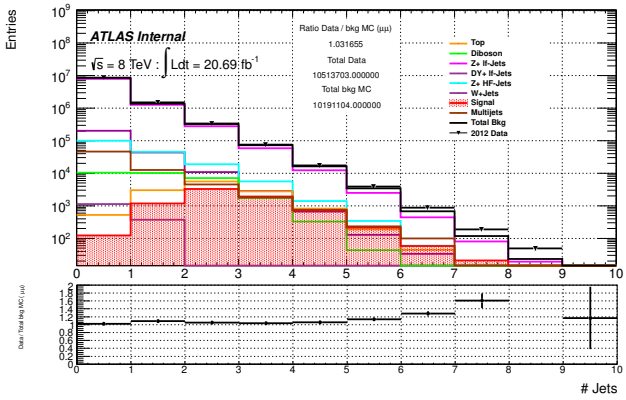
Figure 5.21 Distribution of the leading jet p_{T} (a), subleading jet p_{T} (b), and jet multiplicity after dilepton selection (c) for the electron channel.



(a)



(b)



(c)

Figure 5.22 Distribution of the leading jet p_{T} (a), subleading jet p_{T} (b), and jet multiplicity after dilepton selection (c) for the muon channel.

5.3.3 Kinematic fit

The selection of the jet-jet pair used to build the final $\ell\ell jj$ system, makes use of a χ^2 minimization.

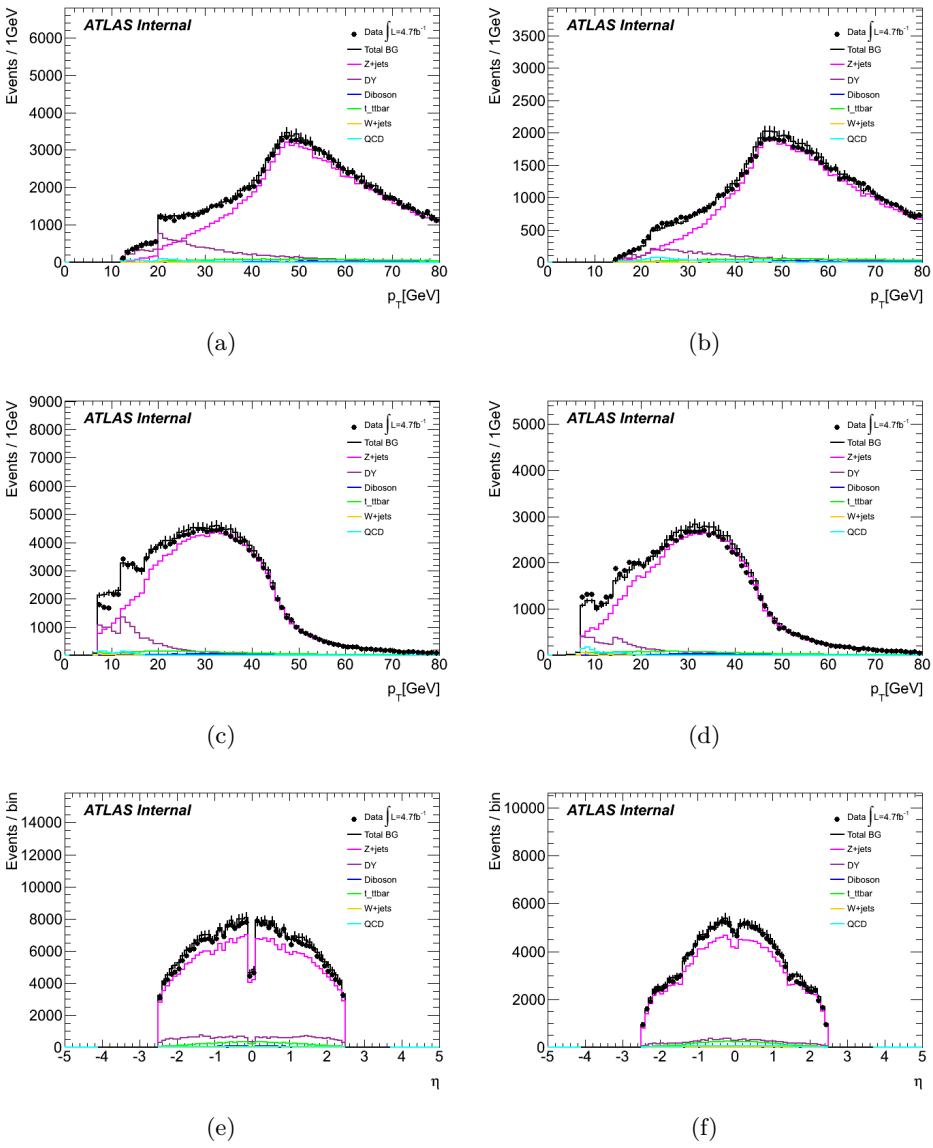
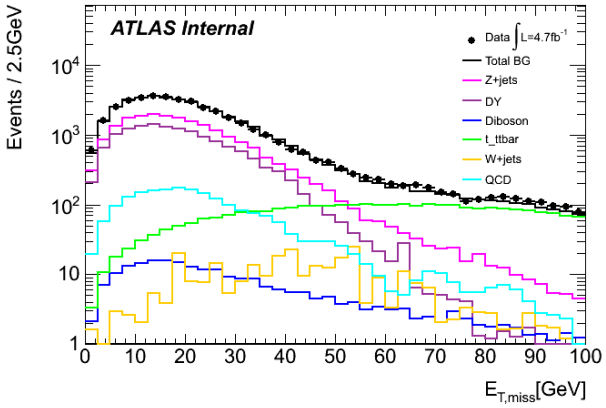
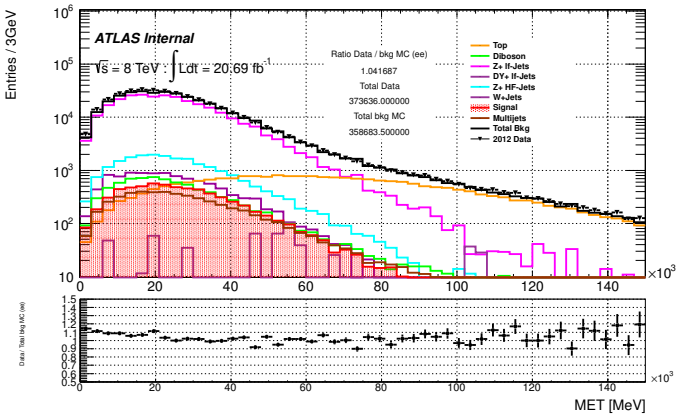


Figure 5.23 The p_{T} of the leading muon (a) and electron (b), subleading muon (c) and electrons (d) and the η distribution of muon candidates (e) and electron candidates (f) after the ≥ 2 jets requirement, for the 2011 Low Mass analysis.

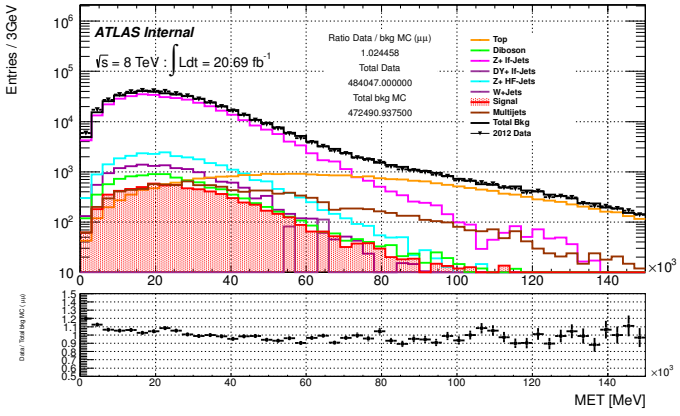
This method aims to select the two jets requiring that the two jets are compatible with the decay of a Z boson. This is true for signal process, while it's not expected for most of the backgrounds (exception is the WZ, ZZ process which is however almost negligible compared to others). The Higgs boson mass resolution is also improved by imposing a Z mass constraint on the invariant mass of the two jets, since the reconstructed $Z \to jj$ mass resolution is much larger than the intrinsic Z width.



(a)



(b)



(c)

Figure 5.24 (a) Distribution of the $E_{\mathrm{T}}^{\mathrm{miss}}$ observed in data compared to MC expectations for electrons and muons together into the 2011 Low Mass analysis. Preliminary $E_{\mathrm{T}}^{\mathrm{miss}}$ distributions for the (b) electron channel and (c) and muon channel in the 2012 High Mass analysis.

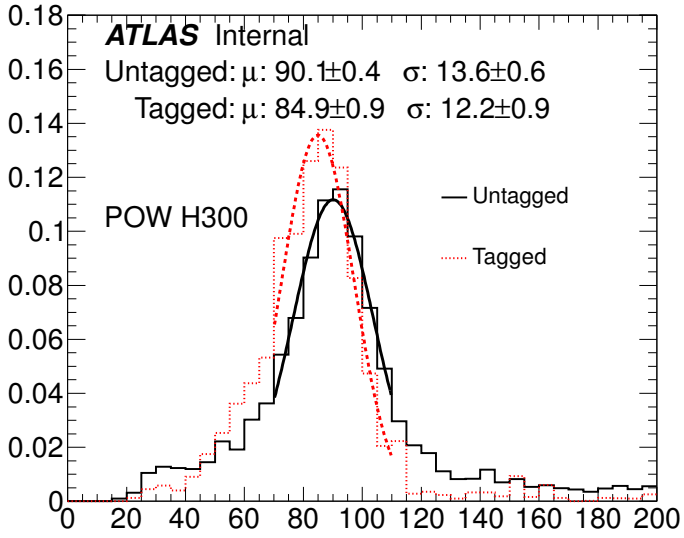


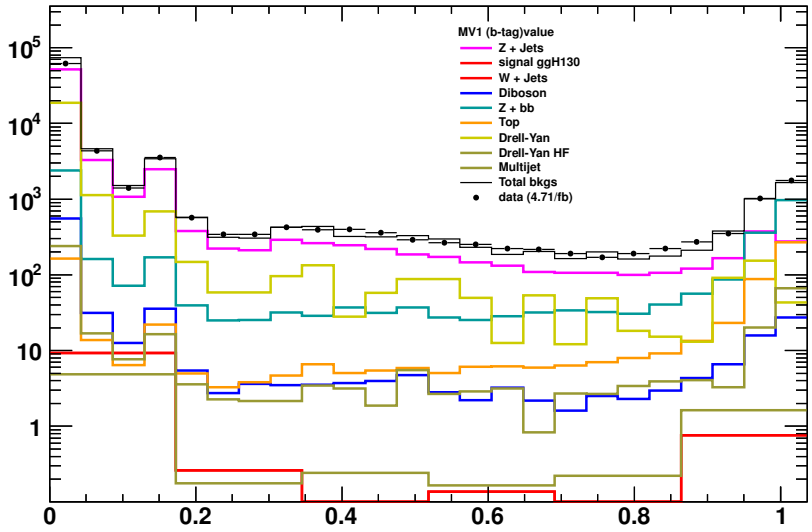
Figure 5.25 Reconstructed mass of the $Z \to b\bar{b}$ process compared to the invariant mass of the $Z \to q\bar{q}$ process. The $Z \to q\bar{q}$ and the $Z \to b\bar{b}$ processes used to build this plot are taken from a sample of Higgs boson with $m_H = 130$ GeV.

The χ^2 is built using jet four-momenta, the expected jet energy resolution σ_{jet_i} measured as a function of the jet η and p_{T} [94] and the world average M_Z and Γ_Z [82]:

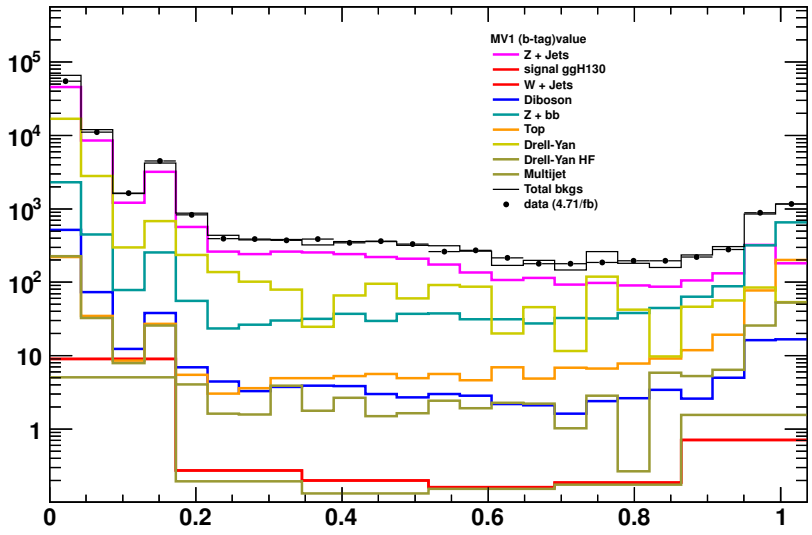
$$\chi^2 = \left(\frac{M_Z - m_{jj}}{\Gamma_Z}\right)^2 + \sum_{i=1,2} \frac{(p_{\text{T}_i} - p_{\text{T}_i}^{fit})^2}{\sigma^2_{\text{jet}_i}} \tag{5.1}$$

The χ^2 function is minimized for each jet pair in the event satisfying the jet selection criteria in Table 5.14, varying the jet p_{T} within the constraint given by the individual jet energy resolution σ_{jet_i}. The jet energy resolution is measured in data with the method described in section 3.6.4 and is parametrized in bins of p_{T} and η.

The jet-jet combination which gives the minimum χ^2 have been selected. Furthermore, for the purpose of the Higgs mass reconstruction, the $m_{\ell\ell jj}$ is calculated using the p_{T} corrected according to the fit result. This method has then a dual purpose: it gives first a criteria for choosing the most likely pair to be produced from the Higgs signal, and it improves significantly the reconstruction of the Higgs mass.



(a)



(b)

Figure 5.26 MV1 discriminant for the leading (a) and subleading jet (b) for the 2011 Low Mass analysis.

When dealing with the untagged channel, this approach allows an efficient reduction of the combinatorial background with respect to the strategy consisting in keeping all jet pairs within some definite mass window. In the tagged channel the two jets are already chosen to be the two b-tagged jets, and hence the kinematic fit has the only effect of improving the Higgs mass resolution.

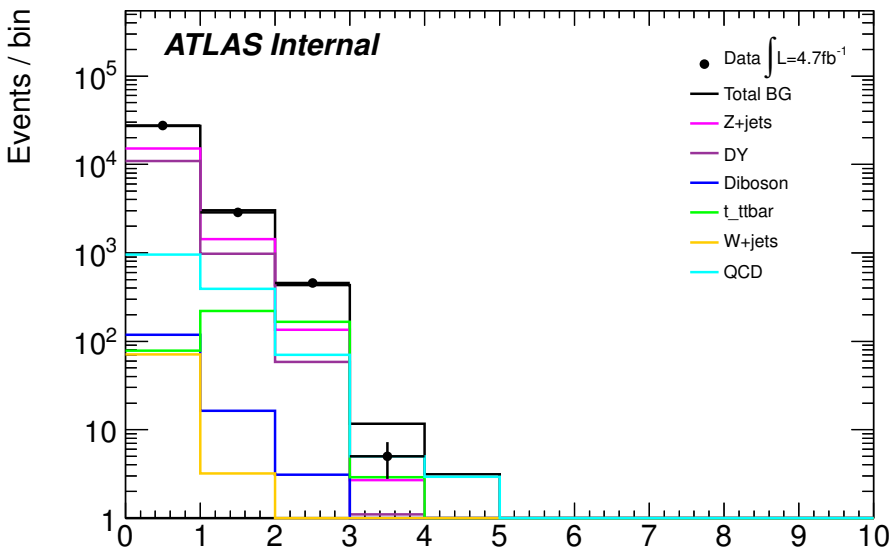


Figure 5.27 Number of b-tagged jets for the muon and electron channel in 2011 Low Mass analysis.

The distribution of the χ^2 after the dilepton plus at least two jets selection is shown in Figure 5.30, while the improvement in the mass resolution for signal events is shown in Figure 5.31 for several mass points. The resolution of the event peaking at the right mass is about 2.7 GeV at 130 GeV Higgs mass. The tail on the high mass side is due to the subset of the event where the dijet system is not coming from an on-shell Z decays. For these events the kinematic fitting is not effective and the resolution on the m_{lljj} variables is unchanged from the raw one.

5.4 Backgrounds in 2011 Low Mass analysis

The main backgrounds for the 2011 Low Mass analysis are discussed in this section. Due to the early state of the 2012 analysis, this will be not introduced in this thesis.

The dominant background is expected to be from Drell-Yan/Z + jets events, with multijet and top production events contributing significantly. Data driven methods are used to determine or control these backgrounds. Smaller contributions are expected from diboson production which are estimated using MC simulation. Finally the small contribution of W+jets events is estimated through simulation, but possible discrepancies are covered in the data driven multijet background estimation.

5.4.1 Drell-Yan/Z + jets background

Drell-Yan (DY) and Z+jets is the dominant background in the untagged case, while the Drell-Yan and Z+ heavy jets process has the higher contribution in the tagged case. ALPGEN MC is used to model them. Since ALPGEN is a LO MC and processes with jets are notoriously difficult to predict, therefore uncertainties on their cross sections are large, and because of this it is convenient to constrain both the overall normalization and the shape of the Drell-Yan and Z + jets using control regions. The control regions for the Drell-Yan/Z + jets background are the side-bands of the m_{jj} window required in the analysis (defined in the previous section), and the check is performed building the final $m_{\ell\ell jj}$ distribution replacing the nominal m_{jj} window cut with the requirements that the dijet mass lies in the side-band region. At the end of the selection the data and MC expectations are compared.

As for the DY/Z+jets, also the normalization of the top background in the tagged case is estimated from a control region (details are given in section 5.4.2). During the procedure to obtain the sample normalization the other backgrounds are subtracted from data, and hence the result found for a given sample (say the DY/Z+jets) may affect the normalization of the other sample (as for top) and viceversa.

To avoid this effect, the normalizations of the DY/Z+jets and top backgrounds are performed in parallel using an iterative estimation. Due to the low available number of events in the tagged case, the electron and the muons channels are considered together. Several systematic checks are performed to verify the robustness of the method and estimate the uncertainties of its application: separate estimates are calculated from the low and the high m_{jj} side bands and the size of the side bands as well as the mass window positions are altered in order to check the stability of the resulting scale factor.

Figure 5.32 presents the comparison between data and simulation for the contribution of the DY/Z + jets background. The obtained scale factors are found to be:

Untagged electron:	$1.02 \pm 0.03 \pm 0.02$
Untagged muon:	$0.99 \pm 0.02 \pm 0.04$
Tagged:	$1.22 \pm 0.13 \pm 0.12$

The above scale factors are obtained as the ratio between the data from which all the background contributions but the DY/Z + jets are subtracted and the DY/Z+jets background itself as it is obtained by the fully MC-based estimation. It is worth to remark that figure 5.32 shows a very good agreement between data and MC. This is confirmed by the scale factors which are consistent with one within the errors. The systematic uncertainty in the untagged case receives contributions both from the multijet subtraction and the variation of the side-band windows. In the tagged case the systematic uncertainty is dominated by the uncertainty on the subtracted contributions of top and multijet backgrounds.

5.4.2 Top background

Top production constitutes a significant background in the tagged case and gives also a small contribution in the untagged case. This background is dominated by leptonic $t\bar{t}$ decays in which the leptons originate either from the W boson decays or the $b-$jets from the top quark decays as shown in chapter 1. The presence of neutrinos in the leptonic decays of $t\bar{t}$ leads to large values of missing transverse energy. Therefore, the requirement of low missing energy (see 5.3.1) reduces this background considerably. The isolation requirement on the leptons reduces further the contribution of leptons originating from hadrons in the top decays.

As for the DY/Z+jets background, also the top normalization is estimated from data, studying the control region obtained by inverting the $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement ($E_{\mathrm{T}}^{\mathrm{miss}} > 40$ GeV). The inverted $E_{\mathrm{T}}^{\mathrm{miss}}$ control region of the tagged case is dominated by top decays, receiving a small contribution (of 4%) from multijet events which is estimated with the *ABCD method* described below.

The comparison between data and simulation for the top background contribution in the tagged sample computed from the inverted $E_{\mathrm{T}}^{\mathrm{miss}}$ control region is presented in Figure 5.33, and the scale factor obtained from this comparison is the following:

$$\text{Tagged:} \quad 1.09 \pm 0.06 \pm 0.04$$

where the first error is statistical and the second systematic. As for the DY/Z+jets background, a good agreement is found between data and purely MC-based estimation. This is confirmed by the obtained scale factor which is found to be consistent with unity. Nevertheless this scale factor is applied to the top quark background estimation for the subsequent studies. The dominant systematic contribution arises from the subtraction of the other processes in the control region. The DY/Z+jets in the untagged case and the multijet background in the tagged case. The calculation of the systematic uncertainty is described in section 5.5.

5.4.3 Diboson background

Even if the Standard Model $ZZ \to \ell^+\ell^- q\bar{q}$ process is the only irreducible background, it turns out to be not so important in this analysis since the huge contribution from the DY/Z+jets production. Hence in this analysis it is completely estimated from the MC simulation. An additional small background from WZ production, which is dominated by the case in which the Z boson decays leptonically and the W boson decays hadronically. This background is also estimated from MC simulation as well as the WW process, which is expected to give an even smaller contribution.

5.4.4 Multijet background

The multijet (QCD) background constitution is significant in the low mass dilepton region where this search is performed, since this background is difficult to model, two distinct data-driven methods where used to evaluate both its size and the shape for the final analysis. Some general remarks valid for both the methods have do be given before going through the details of the estimation of this background: in general many QCD-enriched samples are available, and therefore the choice of the default one may be an issue. To overcome this problem, quality checks have been performed on the several samples and the ones used as default resulted to be the best ones in terms of statistical tests performed using the two methods.

It is worth to remark that the choice of the default QCD samples performed independently for the two methods gave the same results for both the muon and electron channel.

Another issue that may affect the multijet background background is the low statistics of the tagged sample, since the QCD templates are taken from data, which usually have much less statistics than the MC samples. For this reason the same shape is assumed for both the tagged and the untagged case, which is the one provided by the final selection without splitting the events in the tagged/untagged categories, while the normalization is estimated separately for the two subsamples, measuring the tagged event rate in the QCD-enriched samples.

Multijet background estimation using a template fit method

The template method aims at taking the multijet background estimation from data. The shape of the background is obtained from a data sample dominated by multijets events and then subsequently normalized to the signal selection. The normalization of this multijet sample is estimated by fitting the dilepton invariant mass spectrum after applying the nominal selection up to the requirement of ≥ 2 jets. This is performed over the mass range $15 < m_{\ell\ell} < 120$ GeV using two components to fit the data distribution:

- The multijet template derived from data using an ad hoc selection (see the following for details in each channel)
- The sum of all the others contributions (including the signal sample) obtained with MC simulations with the nominal selection up to the ≥ 2 jets requirement.

The result of the fit is a scale factor that has to be applied to the multijet sample in order to let it have the right normalization with respect to the data sample. After the mentioned fit is performed, the quality of the result can be checked looking at the obtained χ^2 in its output. The default QCD template is chosen among the available ones as the one for which the template fit gives the best χ^2.

Application of the template fit method for the dimuon sample The data sample used for the template fit method in the muon channel are the following:

- same sign muon pairs, with all the isolation and quality requirements were left unchanged
- opposite-sign muon pairs, inverting the isolation requirement on at least one of the two

The first sample is used as default since it gives a better χ^2 when used for the fit, while the second is used for the systematic studies on both normalization and shape of the QCD background shown in the following. The fit is performed on the dimuon mass distribution obtained after the missing energy and the two jets requirements. The result extrapolated to the dilepton signal region ($20 < m_{\ell\ell} < 76$ GeV) corresponds to a fraction of QCD multijet events with respect to the selected number of data events for both tagged and untagged samples. The normalization of the tagged sample is obtained adding the request of exactly two tagged jets. The results obtained are the following:

$$\begin{aligned} \text{untagged:} \quad & [4.3 \pm 1.0 \text{ (stat)}]\% \\ \text{tagged:} \quad & [10.9 \pm 2.8 \text{ (stat)}]\% \end{aligned}$$

The considerably higher QCD multijet background in the tagged sample is attributed by an increased proportion of heavy flavour decays in the muon multijet sample naturally leading to an higher tagging rate. In both the samples the systematic contribution to the uncertainty due to the variation of the sample is negligible.

Besides the normalization it is important to establish also the expected $m_{\ell\ell jj}$ shape for this background. To do this the two multijet-enriched selections have been compared and the impact on the $m_{\ell\ell jj}$ modelling of either of the choices (see Figure 5.34(a)).

Application of the template fit method for the di-electron sample The multijet samples used in the dielectron channel can be obtained varying the sign and the isolation of the pairs (as in the muon channel), but in the electron channel also the quality of the reconstructed electrons can be varied. The possible QCD-enriched samples used in this study are:

- Tight++, opposite sign pairs, in which exactly one of the two electrons is required to be isolated and the other one anti-isolated[5].

[5]We define anti-isolation as the inverse cut in the Track isolation described in Tables 5.11, 5.12 and 5.13.

- Tight++, same sign pairs, one isolated electron and the other anti-isolated
- Medium++ (excluding Tight++), opposite sign pairs with the same isolation requirements as above
- Loose++ (excluding Medium++), opposite sign with the same isolation requirements as above

Among the above samples, the first one is used as default, while the others are used for systematic variations. The resulting QCD fraction in the dilepton signal region is:

$$\begin{aligned} \text{untagged:} \quad & [12.0 \pm 1.4 \text{ (stat)} \ \pm 2.3 \text{ (syst)}]\% \\ \text{tagged:} \quad & [11.9 \pm 2.9 \text{ (stat)} \ \pm 1.6 \text{ (syst)}]\% \end{aligned}$$

The systematic uncertainty is quoted from the maximum difference of the fit result obtained using the other templates described above. The dilepton mass distribution after the template fit is shown for electron and muons in Figure 5.35 for the 2011 Low mass analysis and the 2012 High mass analysis.

As for the dimuon cases a comparison of the $m_{\ell\ell jj}$ shape obtained with different selections aimed at enhancing the contribution of QCD multijet fakes is shown in Figure 5.34(b).

Background estimation using an ABCD method

The other procedure that is used to estimate the QCD background is the ABCD method, based on modified charge and isolation criteria for the selected leptons. The four regions to be defined:

A: events with leptons of opposite charge, both isolated
B: events with leptons of opposite charge, one isolated one non isolated
C: events with leptons of same charge, both isolated
D: events with leptons of same charge, one isolated one non isolated

Where A is the "signal region" and regions B, C and D are dominated by multijet processes. Considering there is no correlation between the charge and isolation requirements, allowing to estimate the expected number of multijet events in signal region A from the number of multijet events in regions B, C and D, assuming the ABCD relation A = B × (C/D) within statistical uncertainty. To perform the calculation, a profile-likelihood approach is used. Denoting the unknown number of multijet events in region A as μ_U, where the number of multijet events in each region can expressed by introducing two nuisance parameters τ_B , τ_C:

A: μ^U
B: $\mu^U \tau_B$
C: $\mu^U \tau_C$
D: $\mu^U \tau_B \tau_C$

The corresponding total events in each of the regions are then:

$$\begin{aligned}\mu^A &= s^A + b^A + \mu^U \\ \mu^B &= s^B + b^B + \mu^U\tau_B \\ \mu^C &= s^C + b^C + \mu^U\tau_C \\ \mu^D &= s^D + b^D + \mu^U\tau_B\tau_C\end{aligned}$$

Where $s^{A,B,C,D}$ and $b^{A,B,C,D}$ are the known contributions from signal and electroweak background processes in each region. The likelihood function is the product of the four likelihoods for the counting experiments in the four regions:

$$L(n_A, n_B, n_C, n_D \mid \mu^U, \tau_B, \tau_C) = \prod_{i=A,B,C,D} \frac{e^{-\mu_i}\mu_i^{n_i}}{n_i!} \tag{5.2}$$

The parameter of interest μ^U and the nuisance parameters τ_B,τ_C are calculated from the minimization of $\log L$. Therefore, the application of the ABCD method provides the normalization of the multijet background in the signal region.

Once the normalization is found, it has to be applied to a template used to model the shape of the QCD background. As said above several templates are available to model the QCD shape in both muon and electron channels. Among the available samples, the default ones in the two channels are chosen as follows: the data-MC agreement in both signal region and m_{jj} sidebands is probed performing Kolmogorov-Smirnov tests. The samples with the best Kolmogorov-Smirnov probability are chosen as default, and the others are used for systematic variations.

Application of the ABCD method for the dimuon sample As described in the previous section, the application of the ABCD method defines the normalization of the multijet background in the signal region. The second ingredient needed for the complete background estimation is the shape of the corresponding background. Also using this approach the samples listed in 5.4.4 are used for default and systematic variations.

The statistical uncertainty on the evaluation of the multijet background is determined by the number of events of the B, C and D regions. Systematic uncertainties include either uncertainties on the normalization or the shape of the distributions. The statistical accuracy of the method is found to be 8% and 40% for the untagged and the tagged case respectively. As a measure of the shape uncertainty, the distributions corresponding to the m_{jj} side bands are used instead of the ones corresponding to the signal region. To estimate the normalization uncertainties, the normalization of the multijet background is estimated at different stages of the event selection. In addition different control regions are studied as a function of the muon "anti-isolation" condition imposed. Adding these contributions in quadrature an estimate of 10% and 34% is found for the untagged and the tagged case respectively.

The final estimate for the QCD multijet background percentage over the total in the dimuon case, is estimated as:

$$\begin{aligned}\text{untagged:} &\quad 3.49 \pm 0.24 \pm 0.35\% \\ \text{tagged:} &\quad 11.0 \pm 4.4 \pm 3.7\%\end{aligned}$$

Application of the ABCD method for the dielectron sample In the electron channel there is one more degree of freedom, both in determining the normalization factor from the ABCD method and also in using a control region of events to provide the shapes. This is the electron identification quality. The best combination is chosen according to the quality of the description of data by the estimation both for the side band region of m_{jj} and the signal region between 100 and 300 GeV in m_{eejj}. Also with this approach the samples used as default and for systematic variations are the same listed in 5.4.4. As said at the beginning of section 5.4.4 because of lack of statistics in the tagged analysis bin, the same shape distribution is assumed for both the tagged and the untagged case, which is the one provided by the final selection without the b-tagging requirement.

The systematic uncertainties are estimated in a similar way as in the dimuon case. The final estimate for the QCD multijet background percentage over the total in the dielectron case, is estimated as:

$$\begin{aligned} \text{untagged:} \quad & 6.57 \pm 0.6 \pm 2.3\% \\ \text{tagged:} \quad & 11.5 \pm 5.1 \pm 2.6\% \end{aligned}$$

The final fraction of QCD background used for the normalization of this background is a combination of the results obtained with the two independent methods described above. In particular the weighted mean of the two estimations is taken as central value and the maximum between the difference of these two values and the single systematics is taken as global systematic uncertainty.

Thus the results are:[6]

channel	central value (%)		stat. unc. (%)		syst. unc (%)
muon untagged:	3.99	±	0.27	±	0.51
muon tagged:	11.0	±	4.4	±	3.7
electron untagged:	9.7	±	0.9	±	3.1
electron tagged:	11.5	±	5.1	±	2.6

5.4.5 Summary of backgrounds

The comparison between data and MC simulation for the m_{jj} side band region, using all the data driven corrections described in the previous sections, is presented in Figure 5.36.

From Figure 5.36 it is possible to see the good agreement between data and MC expectations over all the $m_{\ell\ell jj}$ spectrum in both tagged and untagged case, for electrons and muons. The statistical uncertainties of the MC templates are shown since after all the selection they appear to be not negligible with respect to the usually dominant statistical uncertainty on data.

[6] Expressed as fraction of multijet events with respect to the data in the final $m_{\ell\ell jj}$ sample.

5.5 Systematic Uncertainties in 2011 data

This last section describes the calculation of the systematic uncertainties for this 2011 Low Mass analysis. Several systematic uncertainties have been considered: an important contribution comes from theoretical uncertainties on the signal cross section. In addition, theoretical uncertainties on the background processes have to be taken into account as well as systematics that may affect the data driven estimations performed for some of the backgrounds, as described in 5.4. Moreover systematics coming from the experimental apparatus and techniques have to be taken into account.

Reconstruction and identification The main detector-related contributions to the systematic uncertainties are the lepton, jet, and $E_{\mathrm{T}}^{\mathrm{miss}}$ reconstruction and identification efficiencies, their momentum or energy resolution and scale, and the b-tagging efficiency and mistagging rates, and can be computed measuring the effect on the final signal, background and data yields after having varied the definition of these objects.

Concerning the muon (electron) p_{T} (E_T) smearing, the computation of the systematic error derives from the fact that the correction itself is evaluated by means of a Gaussian function using its central value for the nominal correction: systematic variation are obtained varying the central value by ± 1 σ. The uncertainties on lepton reconstruction, identification and trigger efficiencies are mainly of statistical nature since they are measured from data with the *tag and probe* method, but they have also systematic uncertainty coming from the definition of the tag and the probe leptons used to perform such a measurement.

Jet-related uncertainties include the jet energy scale and resolution uncertainties, which have a direct impact on the signal selection efficiency. Those include uncertainties for close-by jets and the fraction of quarks and gluons in the sample. For b-jets an extra scale uncertainty of between 1% and 2.5%, depending on the jet p_T, is added in quadrature in order to take into account the dependence of the jet energy scale and resolution on the jet flavour. Uncertainties on the MV1 b-tagging efficiency and mistag efficiency are evaluated taking into account the tagger working point, and the fraction of the various quark flavour in the MC samples. The uncertainty on $E_{\mathrm{T}}^{\mathrm{miss}}$ is obtained by propagating the uncertainties on the individual objects as described in 5.2.4. The size of these detector-related uncertainties are summarized in Table 5.19.

The detailed results of the application of these experimental uncertainties on the selection of signal and background in the different branches of the analysis are described below: Figure 5.37 and 5.38 show the effect of the jet energy scale systematic uncertainty on the background and signal, respectively. This is the uncertainty which is expected to have the maximum effect on the shape of the $m_{\ell\ell jj}$ distribution, while it can be seen that also due to the use of the kinematic fit on the dijet system, the effect of this uncertainty on the shape has a reduced impact on both the signal and background shape. The shape uncertainty from jet-energy-scale systematic uncertainty will be neglected.

Source of uncertainty	Treatment in analysis
Luminosity	3.9%
Jet energy scale (JES)	2–7%, as a function of p_{T} and η
Jet pileup uncertainty	3–7%, as a function of p_{T} and η
b-quark energy scale	2.5–1% as a function of p_{T}
Jet energy resolution	1–4%
Electron selection efficiency	0.7–3%, as a function of p_{T};
	0.4–6%, as a function of η
Electron reconstruction efficiency	0.7–1.8%, as a function of η
Electron energy scale	0.1–6%, as a function of η, pileup, material effects, etc.
Electron energy resolution	Sampling term 20%;
	a small constant term has a large variation with η
Muon selection efficiency	0.2–3%, as a function of p_{T}
Muon trigger efficiency	$< 1\%$
Muon momentum scale	2–16%, as a function of η
Muon momentum resolution	p_{T} and η dependent resolution smearing functions,
	systematic $\leq 1\%$
b-tagging efficiency	5–15%, as a function of p_{T}
b-tagging mistag rate	10-22%, as a function of p_{T} and η
Missing transverse energy	Add/subtract object uncertainties in $E_{\mathrm{T}}^{\mathrm{miss}}$
	+ uncertainty on"SoftJet" and "CellOut" terms

Table 5.19 Systematic uncertainties related to object reconstruction and identification

Signal cross sections Higgs boson production cross sections have been studied extensively by the LHC Higgs cross section working group and the results are compiled in [106]. Theoretical uncertainties on the cross sections have been estimated to be between $15-20\%$ for $gg \to H$ and $3-9\%$ for $qq \to qqH$ (VBF) for m_H relevant for this analysis. In addition to this another systematic effect has to be considered in order to take into account the accuracy with which the QCD scale is known. This uncertainty amounts to $\sim 14\%$ for all the Higgs mass spectrum.

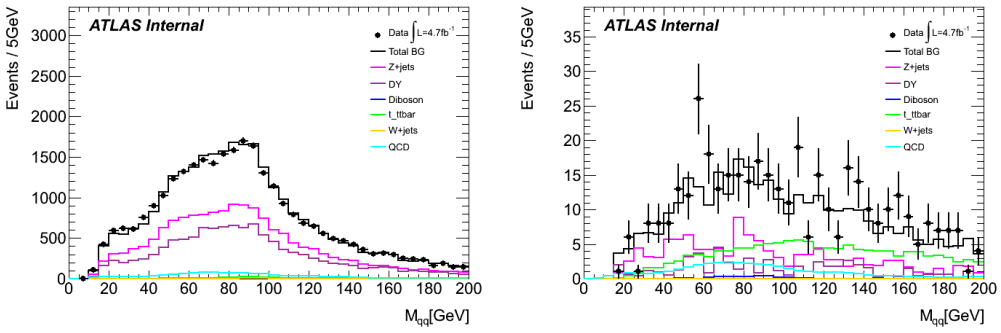
Background normalization The normalization uncertainties of the Z/DY +jets and top backgrounds are estimated through the data driven methods varying the control regions discussed in 5.4, and the variation of the resulting normalization is used as systematic error. The resulting uncertainty for the $\mathrm{DY}/Z+$ jets sample is approximately $3-4\%$ in the untagged channel, while in the tagged case it reaches the value of 17%. The top background uncertainty is estimated to be 10% (7%) for the untagged (tagged) case. The diboson background contributions are assigned a theory uncertainty of 11%.

DY + jets shape As it is shown in Figure 5.36, the shape of the background after normalization corrections is reasonably described by the MC simulation. Nevertheless, the shape uncertainty for the $DY+$jets background is estimated by parametrizing the remaining difference of the $m_{\ell\ell jj}$ distribution of the m_{jj} sidebands, after the application of the normalization scale factor as a function of $m_{\ell\ell jj}$. In the tagged case the statistical accuracy obtained in this sample is not adequate for such a study and therefore the results of the untagged case are used.

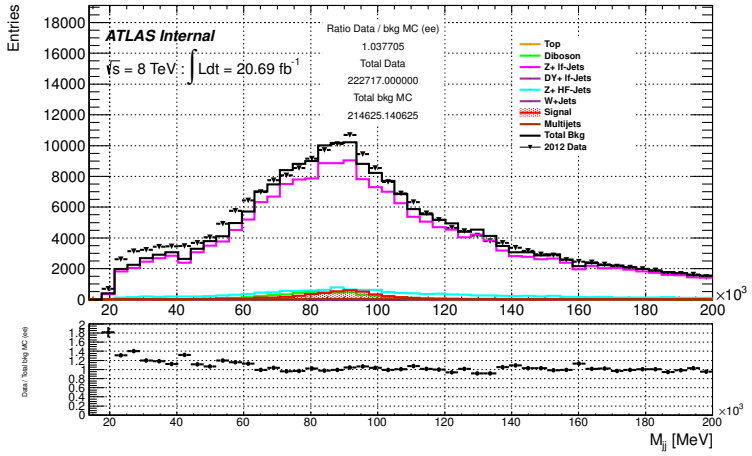
Multijet background normalization The multijet background normalization uncertainty is calculated as described in Section 5.4.4: since the normalization of this background can is calculated with two independent methods, and in each method several QCD-enriched samples can be used, the maximum between the half difference of the two estimations and each systematic (calculated varying the QCD sample) is taken as systematic. The result of this study is that the QCD normalization uncertainty in the muon untagged channel is $\sim 15\%$ in the muon untagged sample, $\sim 35\%$ in the electron untagged channel, while it is $\sim 50\%$ in the untagged channels (both muons and electrons).

Multijet background shape The shape uncertainty on the multijet background is evaluated building the final $m_{\ell\ell jj}$ distribution with the other available multijet samples besides the default ones, as explained in 5.4.4 and shown in Figure 5.34. The shape uncertainty on this background is evaluated to be negligible.

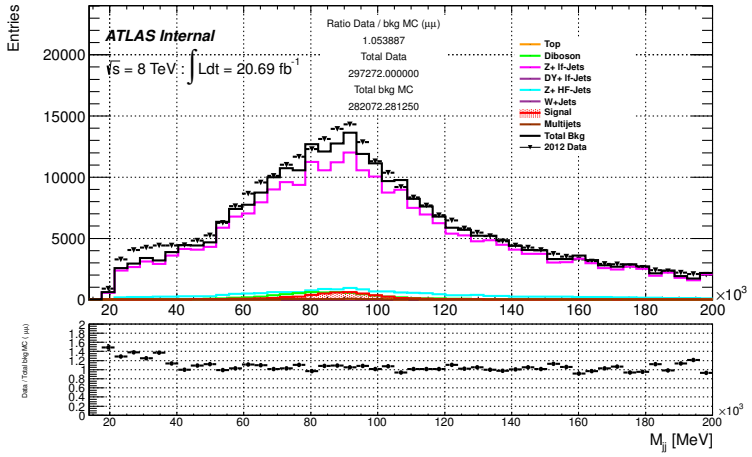
Luminosity The luminosity uncertainty for 2011 data is 3.9%. This uncertainty is only applied to MC samples for which the normalization uncertainty is not taken directly from a comparison between data and MC, which is everything except the DY/Z + jets, the top and the multijet background. Where it is applied this systematic uncertainty is assumed to be correlated across all samples.



(a) (b)

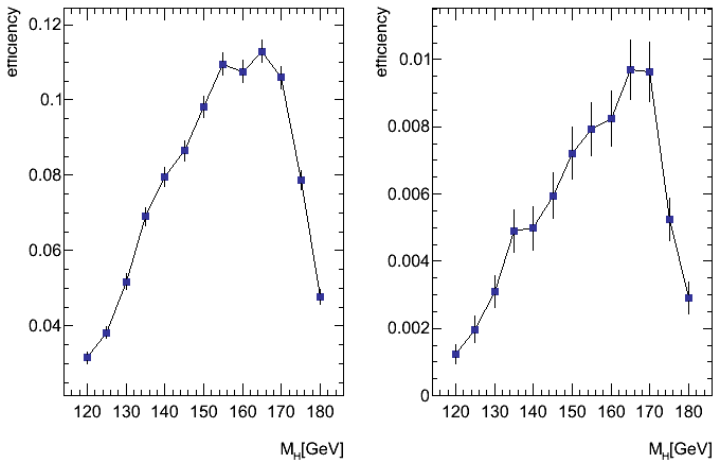


(c)

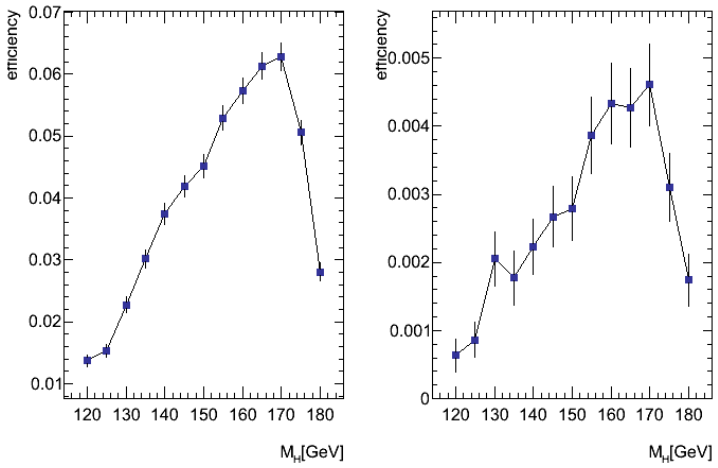


(d)

Figure 5.28 Distributions of the dijet mass after $E_{\mathrm{T}}^{\mathrm{miss}}$ requirement for the untagged (a) and tagged (b) 2011 Low Mass selection for the combined electron and muon sample. (c), (d) untagged m_{jj} distributions for the 2012 High Masss selection (electron and muon channel).



(a)



(b)

Figure 5.29 Signal selection efficiency for the 2011 Low Mass analysis in the muon channel (a) for the untagged (left) and tagged (right) selections and for the electron channel (b).

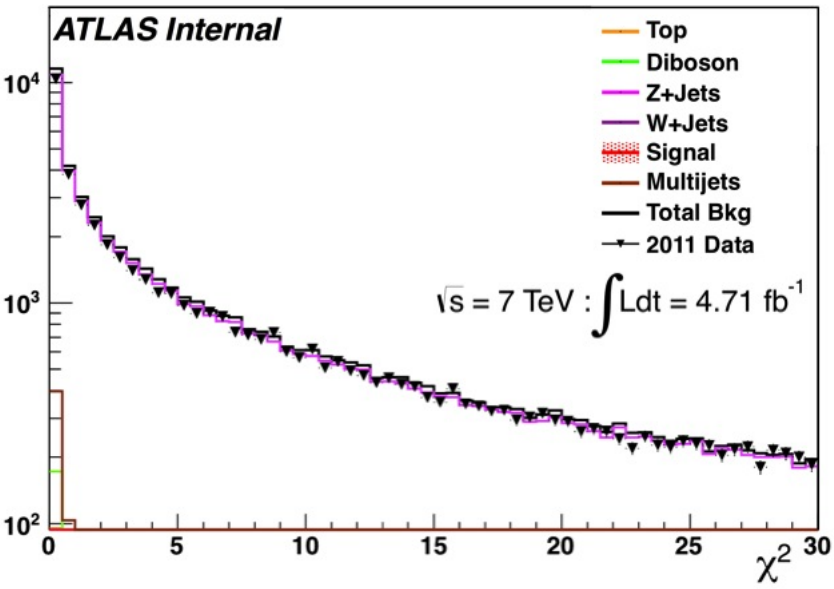
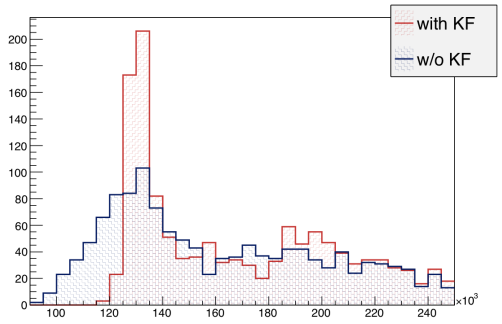
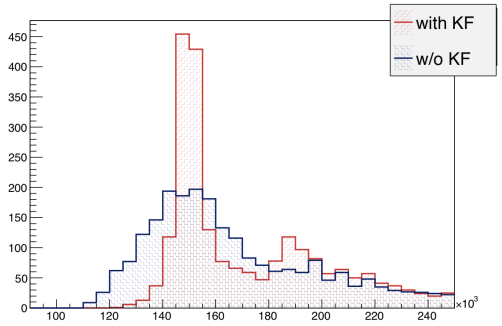


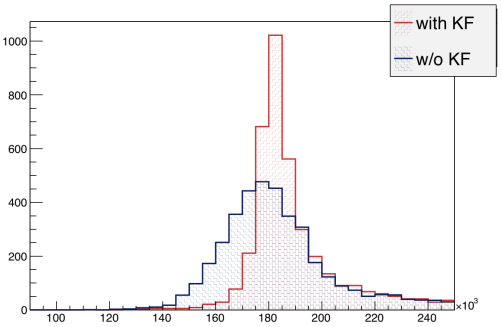
Figure 5.30 The distribution of the χ^2 of the kinematic fitter for Low mass electron channel.



(a)

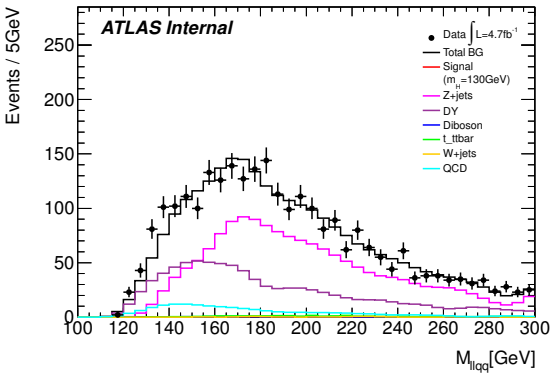


(b)

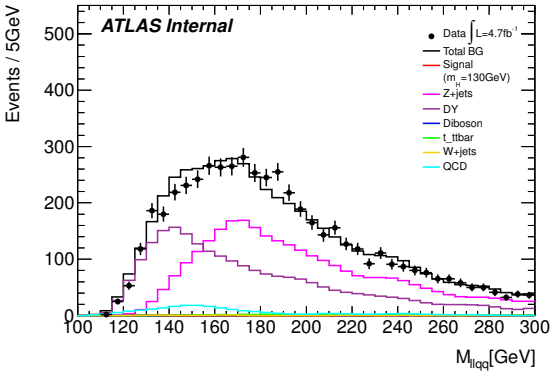


(c)

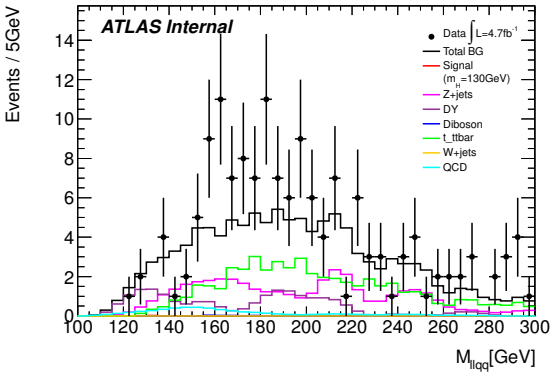
Figure 5.31 The distribution of $m_{\ell\ell jj}$ on signal MC samples before and after the kinematic fitting for three Higgs mass hypothesis: 130 (a), 150 (b) and 180 GeV (c).



(a)



(b)



(c)

Figure 5.32 Comparison between data and simulation for the contribution of DY+jets background, estimated from the m_{jj} side bands, before the application of the scale factor. (a) and (b) correspond to the electron and muon channel respectively of the untagged case. (c) Tagged case for both channels together.

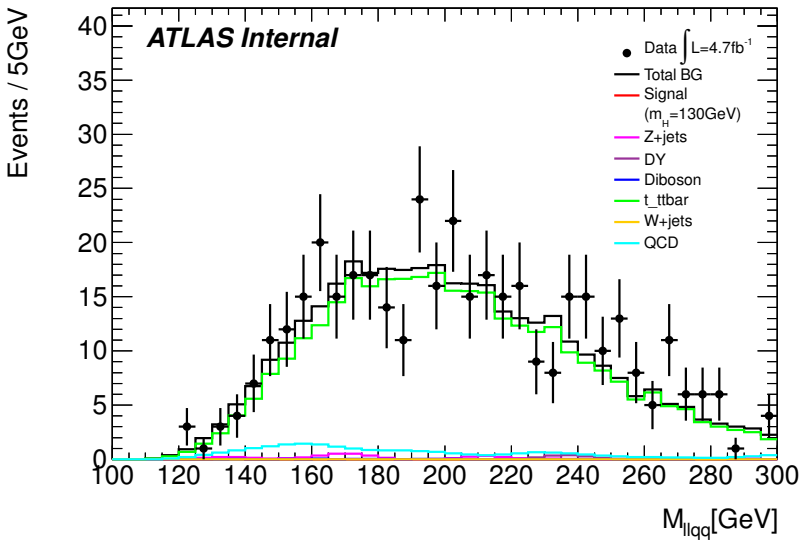
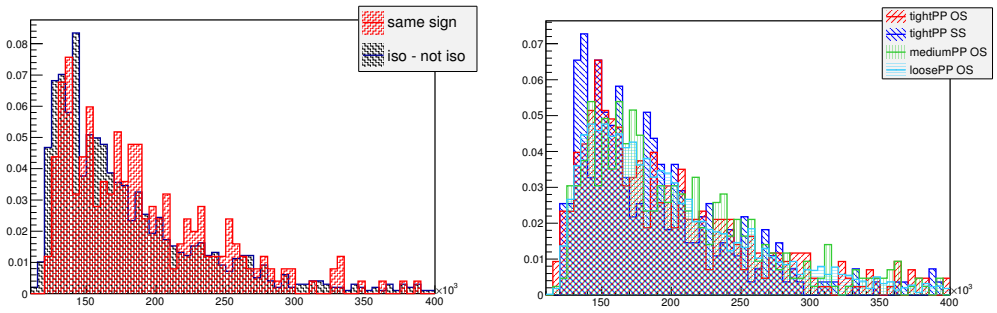
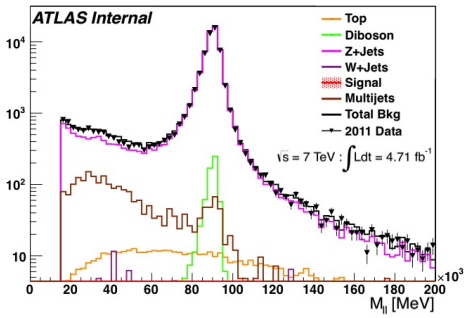


Figure 5.33 Comparison between data and simulation for the contribution of top background in the tagged sample, estimated from the inverted $E_{\mathrm{T}}^{\mathrm{miss}}$ control region, before the application of the scale factor.

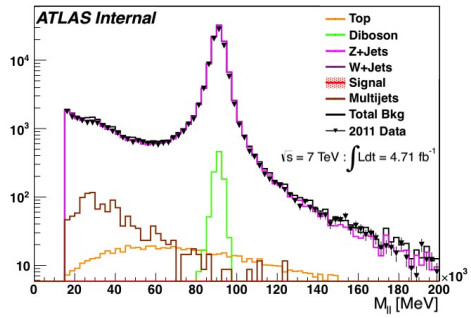


(a) (b)

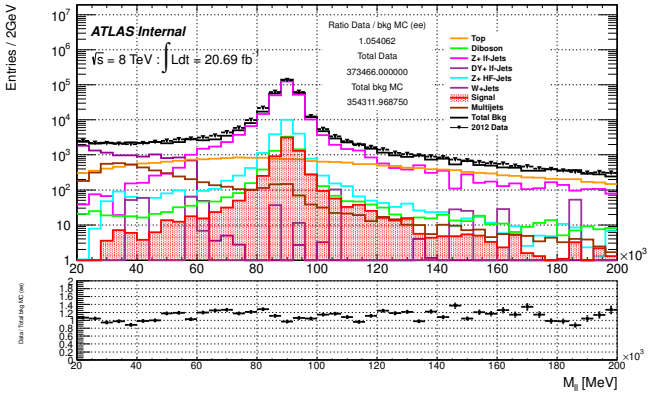
Figure 5.34 Comparison of the $m_{\ell\ell jj}$ distribution for two different selection of QCD dominated control regions for muon sample (a), and electron sample (b).



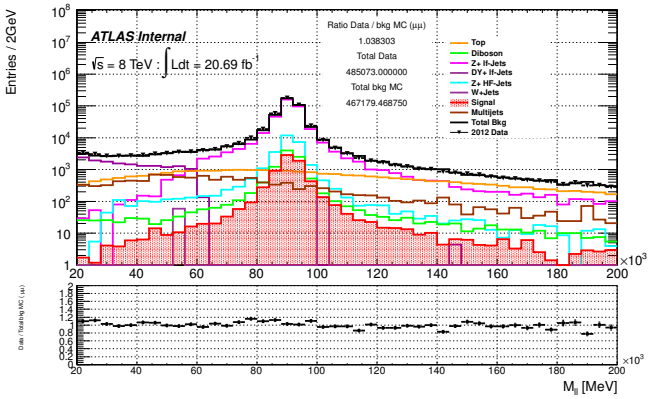
(a)



(b)



(c)



(d)

Figure 5.35 The distributions of the dilepton mass used for QCD fit: (a) for electrons and (b) for muons into the 2011 Low mass analysis and (c) for electrons and (d) for muons the 2012 High mass analysis.

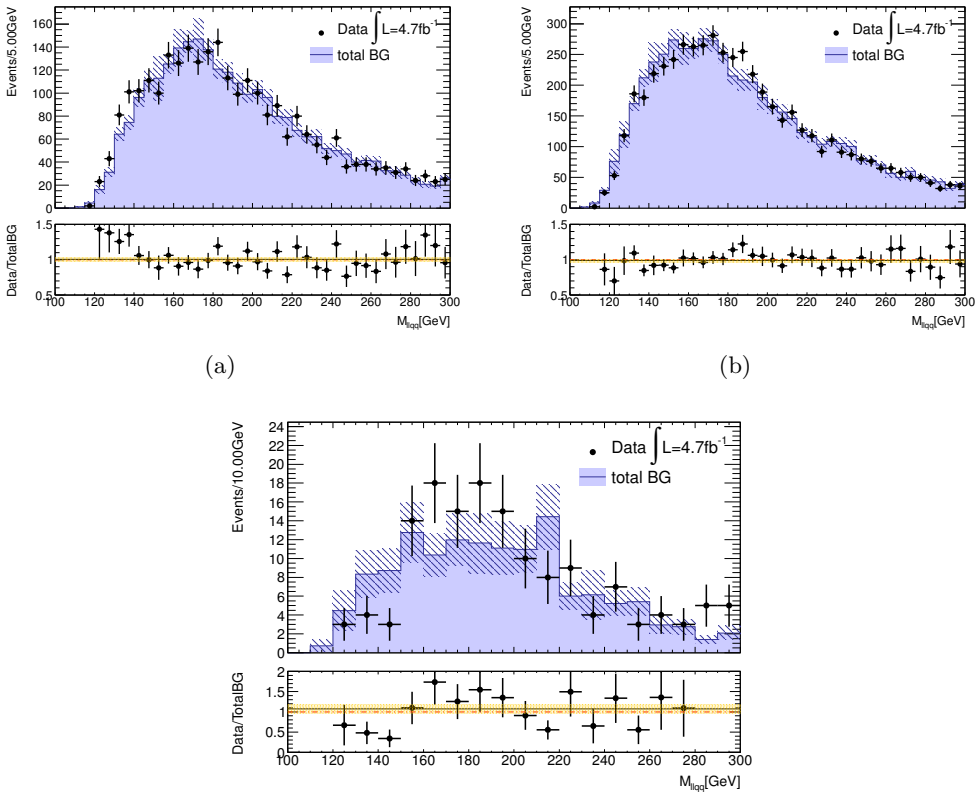


Figure 5.36 Comparison between data and simulation for the m_{jj} side band region, after the application of the scale factors and using the systematic uncertainty estimates. (a) and (b) correspond to the electron and muon channel respectively of the untagged case and (c) to the tagged case for both channels together.

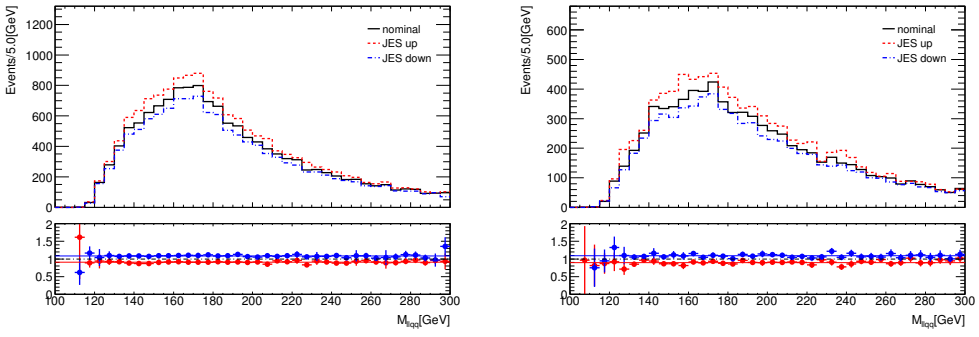
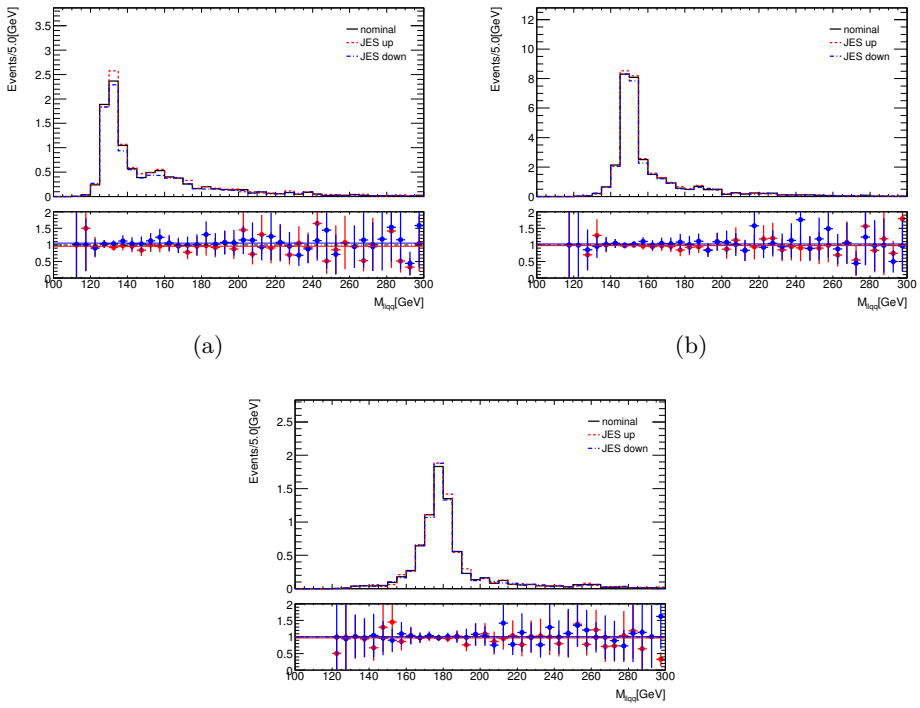


Figure 5.37 Reconstructed Higgs mass for nominal jet energy scale compared to the $\pm 1\sigma$ variation, for background events in the signal region (a) and sideband region (b).



(a) (b)

(c)

Figure 5.38 Reconstructed Higgs mass for nominal jet energy scale compared to the $\pm 1\sigma$ variation, for Higgs signal simulation for 130 (a), 150 (b), and 180 GeV Higgs mass (c).

Chapter 6

Results

Into this chapter the main results of the dissertation will be described, using as references the all previous chapters, stating with the basic theoretical motivations, passing trough the description of the ATLAS detector and the principal objects and variable's reconstruction and Higgs boson searches until the analysis procedure followed. Main results correspond to the 2011 Low Mass analysis, adding a brief status of the Low and High Mass analyses using the 2012 data samples. The search for the Higgs boson is performed by comparing the invariant mass of the $\ell\ell jj$ system, i.e. the reconstructed Higgs boson mass, in the data samples to that of the expected background.

After the kinematic fit (see Section 5.3.3) the expected resolution for the 2011 (130 GeV Higgs mass hipothesis) core signal event distribution is expected to be around 3 GeV. A long tail is also present above the nominal Higgs mass.

The background invariant mass distribution has instead a very broad distribution peaking at around 170 GeV, with a width of around 40 GeV. The distribution of the $\ell\ell jj$ system invariant mass, $m_{\ell\ell jj}$, for 2011 data compare to the predicted background after all data-driven scale factor are applied are shown in Figure 6.2, and Figure 6.3 for the muon and electron channels, respectively.

The expected Higgs signal for 130 GeV mass (multiplied by a factor 20 in the untagged case and by a factor of five in the tagged case) is displayed on top of the background prediction.

A detailed breakdown of the different predicted background sources, data counts and expected signal number of events are also reported in Tables 6.1–6.4, for the four independent analysis channels.

Figure 6.1 described a simplified view of how to interpret the results relative to the variable m_{lljj}.

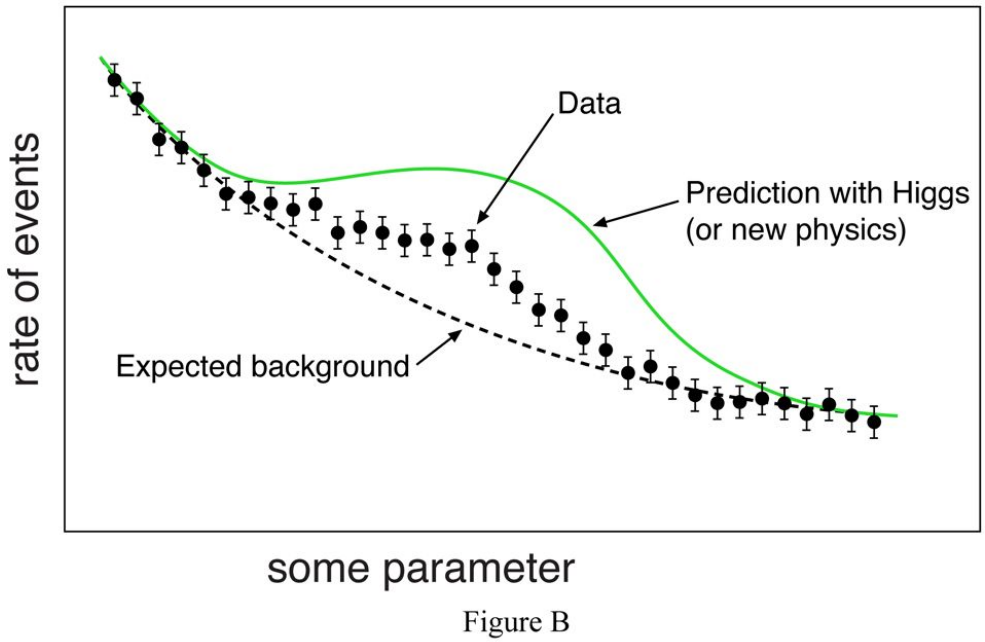
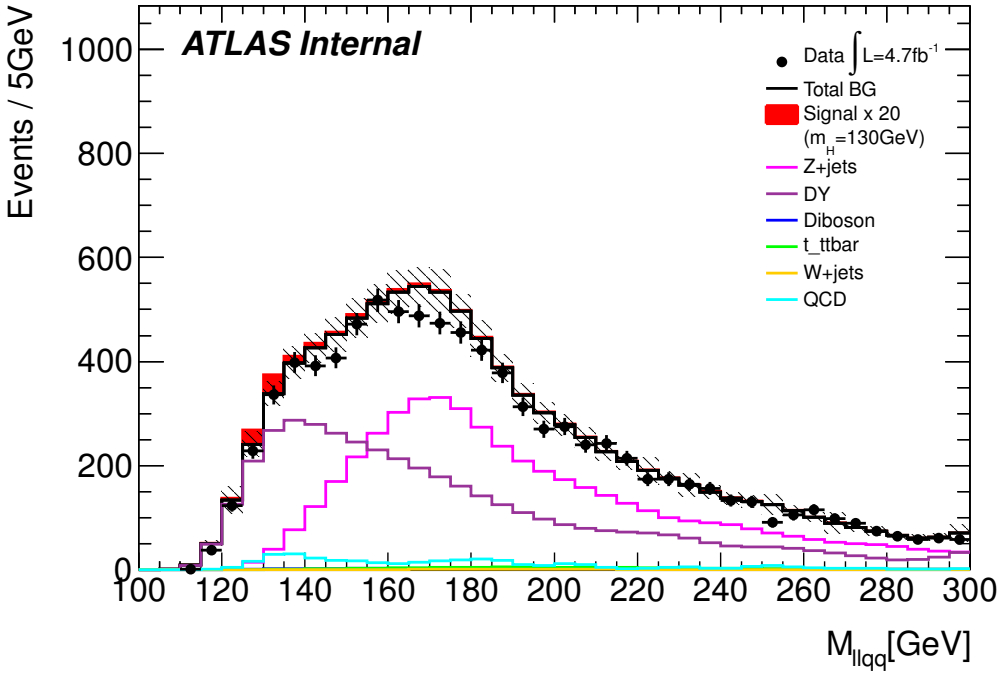
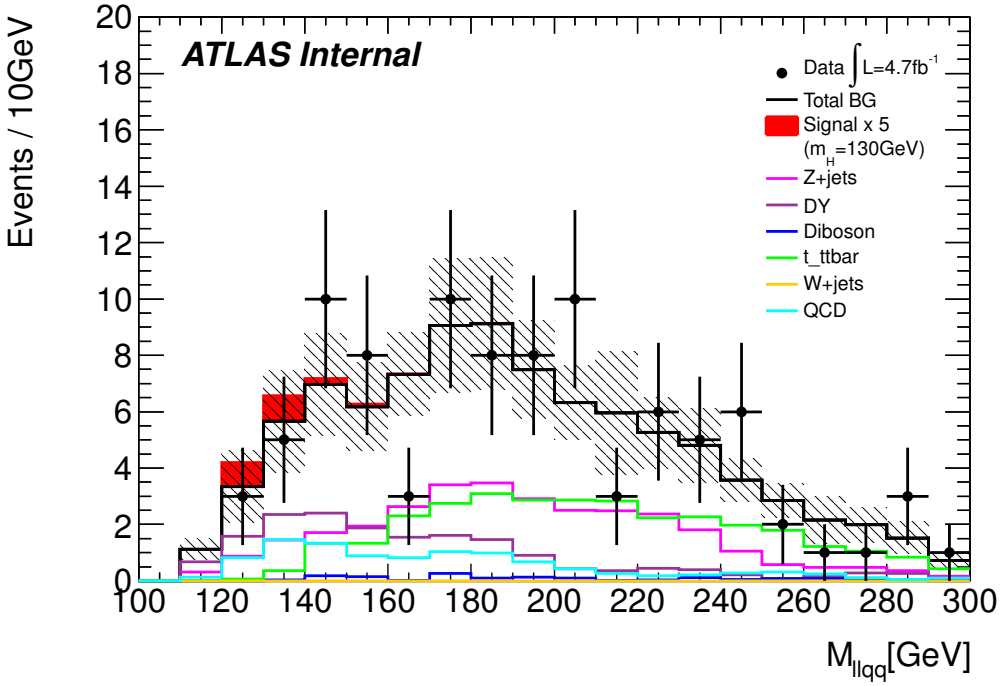


Figure 6.1 Explanatory figure 1: This plot shows hypothetical data and expectations that could be used in setting the CLs. The green curve shows (fictional) predicted results if there were a Higgs boson in addition to all the usual backgrounds. It could also represent the predictions of some other new physics. The dashed black curve shows what is expected from all background processes without a Higgs or some new physics. The black points show the hypothetical data.



(a)



(b)

Figure 6.2 Distribution of the reconstructed Higgs mass in the muon channel with data compared to the background prediction including signal for an Higgs mass hypothesis of 130 GeV. Untag selection (a); Tagged selection (b).

Source	Expected number				
DY+jets	9635	±	101	±	409
Top	99.0	±	1.8	±	9.8
Multijet	355	±	24	±	213
Diboson	60.9	±	1.2	±	9.1
W+jet	10.9	±	2.5	±	1.6
Total background	10161	±	105	±	461
Data	9714				
Signal $m_H = 120$ GeV	2.07	±	0.10	±	0.12
Signal $m_H = 130$ GeV	7.26	±	0.28	±	0.38
Signal $m_H = 150$ GeV	21.1	±	0.60	±	0.73
Signal $m_H = 180$ GeV	4.86	±	0.20	±	0.17

Table 6.1 Summary of numbers of the estimated background events, observed events in data, and expected signal events for the untagged selection in the muon sample. The statistical and systematic uncertainties on the estimated background and signal events are also shown.

Source	Expected number				
DY+jets	53.0	±	5.3	±	7.7
Top	33.8	±	1.0	±	2.2
Multijet	11.0	±	4.4	±	1.6
Diboson	1.7	±	0.2	±	0.3
Total background	99.5	±	7.0	±	8.2
Data	105				
Signal $m_H = 120$ GeV	0.080	±	0.018	±	0.010
Signal $m_H = 130$ GeV	0.431	±	0.067	±	0.051
Signal $m_H = 150$ GeV	1.561	±	0.167	±	0.178
Signal $m_H = 180$ GeV	0.299	±	0.049	±	0.034

Table 6.2 Summary of numbers of the estimated background events, observed events in data, and expected signal events for the tagged selection in the muon sample. The statistical and systematic uncertainties on the estimated background and signal events are also shown.

Source	Expected number				
DY+jets	4654	±	42	±	161
Top	69.0	±	1.5	±	7.8
Multijet	337	±	31	±	450
Diboson	36.3	±	1.0	±	5.3
W+jet	30.2	±	10.7	±	4.3
Total background	5126	±	53	±	478
Data	5197				
Signal $m_H = 120$ GeV	0.90	±	0.06	±	0.07
Signal $m_H = 130$ GeV	3.19	±	0.18	±	0.24
Signal $m_H = 150$ GeV	9.70	±	0.40	±	0.55
Signal $m_H = 180$ GeV	2.85	±	0.15	±	0.14

Table 6.3 Summary of numbers of the estimated background events, observed events in data, and expected signal events for the untagged selection in the electron sample. The statistical and systematic uncertainties on the estimated background and signal events are also shown.

Source	Expected number				
DY+jets	30.5	±	3.5	±	4.4
Top	22.2	±	0.8	±	1.6
Multijet	7.0	±	3.2	±	2.5
Diboson	1.1	±	0.2	±	0.2
Total background	60.8	±	4.8	±	5.3
Data	51				
Signal $m_H = 120$ GeV	0.042	±	0.016	±	0.006
Signal $m_H = 130$ GeV	0.288	±	0.055	±	0.037
Signal $m_H = 150$ GeV	0.594	±	0.100	±	0.072
Signal $m_H = 180$ GeV	0.177	±	0.039	±	0.021

Table 6.4 Summary of numbers of the estimated background events, observed events in data, and expected signal events for the tagged selection in the electron sample. The statistical and systematic uncertainties on the estimated background and signal events are also shown.

6.1 Exclusion confidence level determination in 2011 Low Mass analysis

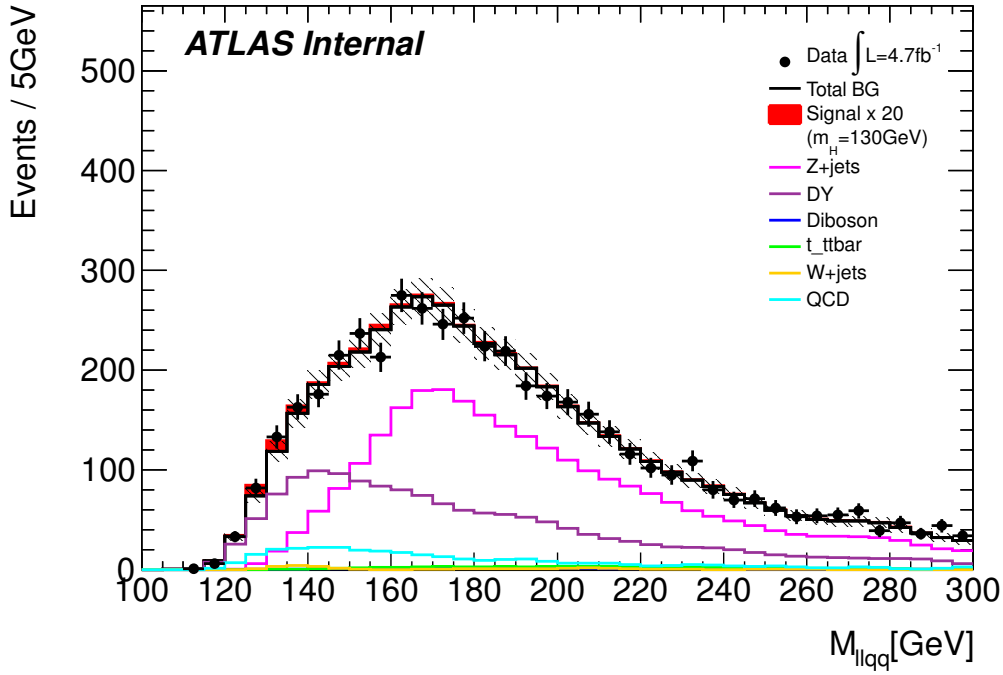
Since the previous results don't show a significant excess on the observed events over the expected background, 95% CL limits are set on the Higgs boson cross section as a function of mass, using the CL_s modified frequentist formalism with the profile likelihood test statistic (see Section 4.3.1)[145, 146]. Figure 6.4 described a simplified view of how to interpret the results relative to the CL.

6.1.1 Results using MCLIMIT software[147]

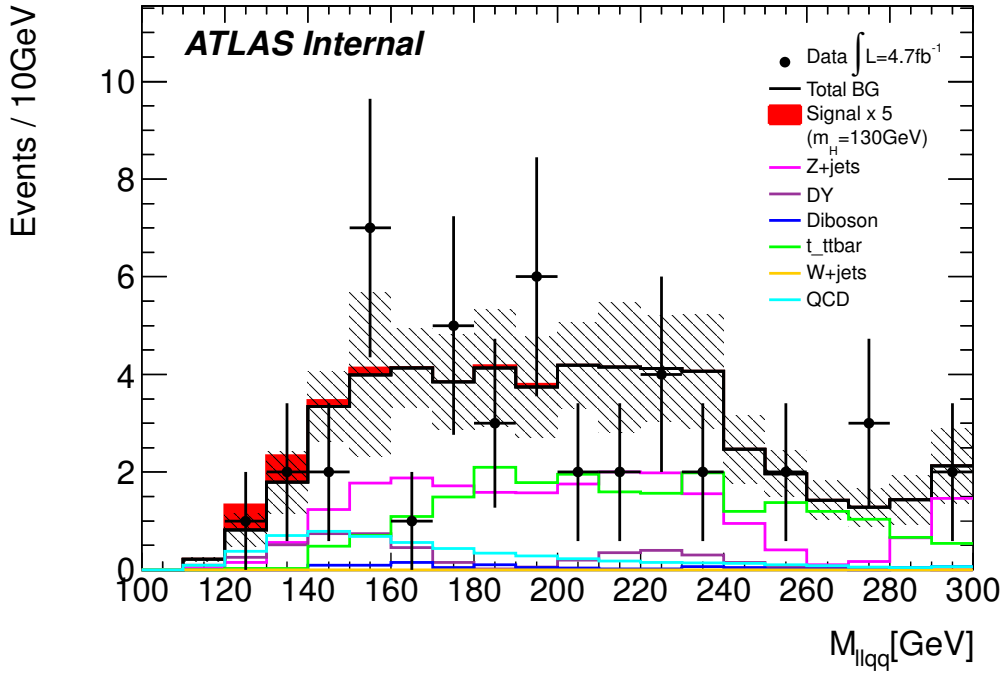
As not significant excess is observed in the data above background expectations, it possible to set limits on the range of possible values of m_H for all studied decay channels using MCLIMIT. The software is based on the mentioned CL_s method and allows multiple channels to be combined while taking into account normalisation and shape uncertainties on signal and backgrounds. In this method the quantity CL_s is defined as $CL_s = CL_{s+b}/CL_b$, where CL_{s+b} is the probability that a hypothesis of signal and background will fluctuate to the observed number of data events or lower and CL_b is the probability that a background only hypothesis will fluctuate to the observed number of data events or higher. Low values of CL_b indicate a possible signal and values of $1 - CL_s > 0.95$ mean that the model can be ruled out at a 95% confidence level. Such a upper limits on Higgs boson cross section as a function of mass has been derived first using the software MCLIMIT examining the $m_{\ell\ell jj}$ distribution for the four independent analysis channels, shown in Figures. 6.2, and Figure 6.3 and the systematic uncertainties described in Section 5.5. The Figure 6.5 shows the resulting exclusion. The ratio of the observed to the expected limit, along with median expected limit and the 1σ and 2σ fluctuations, for the m_H values considered are given in Table 6.5.

m_H	Observed	Expected $\mu/\mu_{\rm SM}$				
(GeV)	$\mu/\mu_{\rm SM}$	-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
mH=120	50.4	7.7	28.1	51.7	88.0	113
mH=125	32.2	3.1	11.3	21.3	38.3	64.0
mH=130	12.9	1.2	4.0	9.5	16.0	22.8
mH=135	7.3	1.0	3.2	5.4	8.7	13.4
mH=140	5.8	1.0	3.2	4.8	7.4	10.5
mH=145	3.7	1.0	2.0	3.6	6.1	8.0
mH=150	5.7	1.0	1.9	4.1	6.3	8.1
mH=155	8.0	0.4	1.9	3.7	5.8	8.0
mH=160	6.6	2.0	4.0	7.8	11.9	15.4
mH=165	4.7	0.8	6.2	13.8	19.0	30.3
mH=170	4.0	1.0	6.8	13.6	23.3	34.8
mH=175	11.5	0.9	8.1	16.0	25.7	32.0
mH=180	19.5	1.0	12.5	19.3	31.2	46.5

Table 6.5 The observed 95% CL upper limits in the $H \to ZZ \to \ell\ell jj$ channel, as a multiple of the Standard Model rate, for an integrated luminosity of 4.71 fb^{-1}. The corresponding expected median upper limit, along with its $\pm 1\sigma$ and $+2\sigma$ values, is also shown. These upper limits were obtained with the MCLIMIT software.



(a)



(b)

Figure 6.3 Distribution of the reconstructed Higgs mass in the electron channel with data compared to the background prediction including signal for an Higgs mass hypothesis of 130 GeV. Untag selection (a); Tagged selection (b).

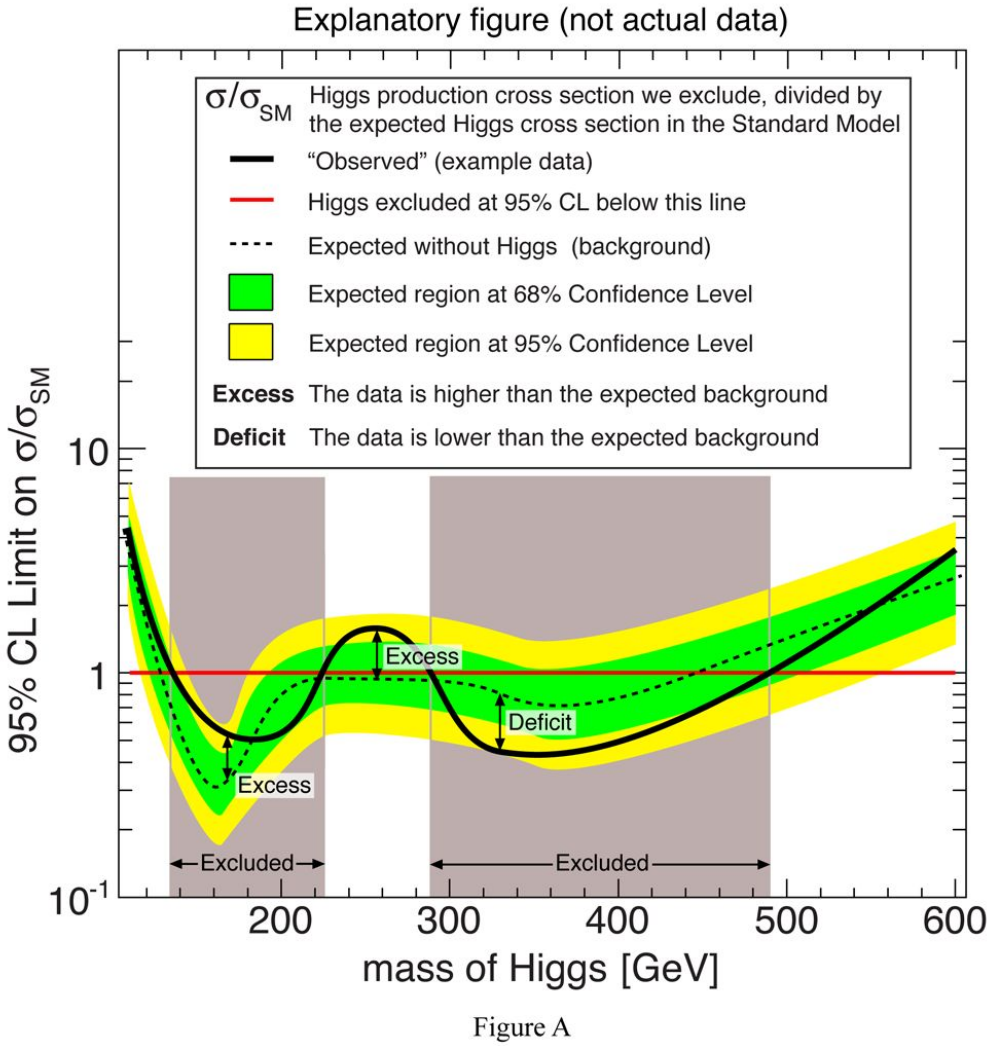
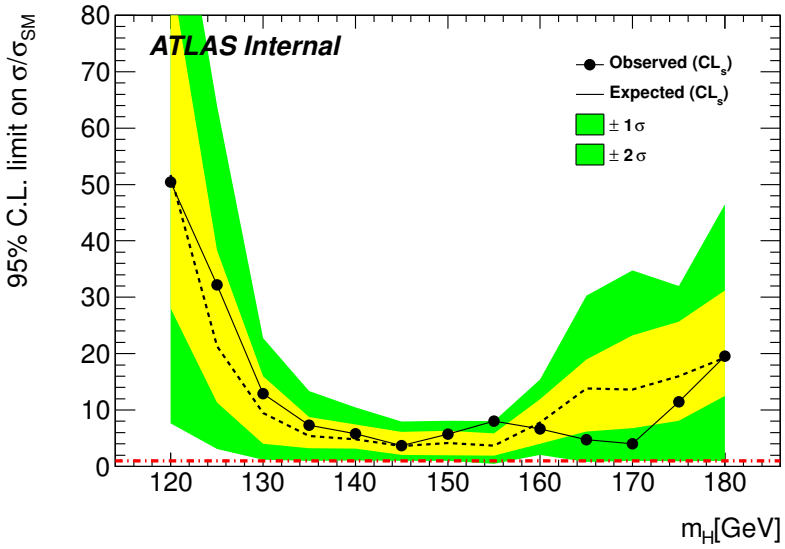
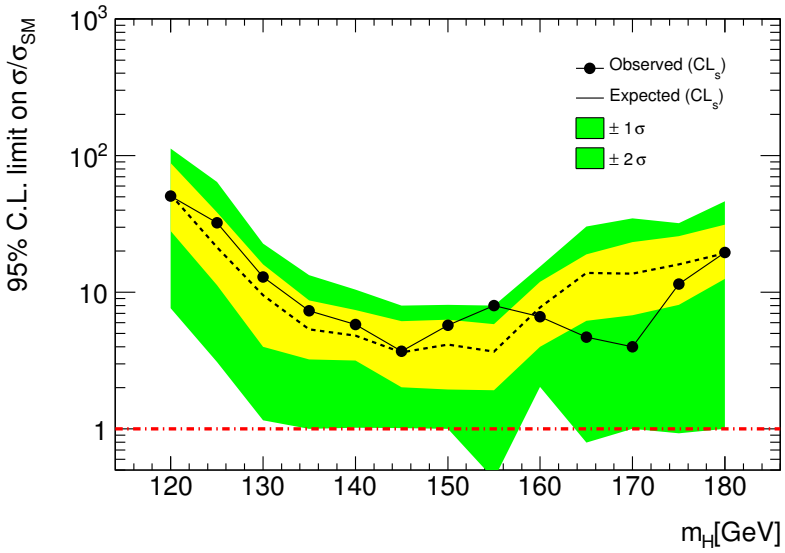


Figure 6.4 Explanatory figure 2: This shows a 95% confidence level, which in effect means the certainty that a Higgs particle with the given mass does not exist. The dotted black line shows the median (average) expected limit in the absence of a Higgs. The green and yellow bands indicate the corresponding 68% and 95% certainty of those values. If the solid black line dips below the value of 1.0 as indicated by the red line, then we see from our data that the Higgs boson is not produced with the expected cross section for that mass. This means that those values of a possible Higgs mass are excluded with a 95% certainty. If the solid black line is above 1.0 and also somewhat above the dotted black line (an excess), then there might be a hint that the Higgs exists with a mass at that value.



(a)



(b)

Figure 6.5 The expected (dashed line) and observed (solid line) upper limits on the total cross section divided by the expected Standard Model Higgs boson cross section, calculated using CL_s at 95%. Linear scale (a), log scale (b). The green and yellow bands, centred on the dotted line, indicate the range in which the limit is expected to lie in the absence of a signal. The dotted red line shows the Standard Model value of 1.

6.1.2 Results using RooStats[148]

For the final results, the official ATLAS code to derive upper limits in the asymptotic approximation have been used, the so-called HistFactory tools[148]. The most notable difference with respect to the result shown in the previous section is the possibility to include the effect of the bin-by-bin statistical uncertainties on background distributions due to limited MC or control region statistics. These are implemented as a set of additional independent nuisance parameters, one for each of the bins included in the analysis, which are allowed to vary within the expected background template statistical uncertainties. A numeric comparison of the result obtained with and without considering this source of uncertainty are presented in Table for the 130 GeV Higgs mass point 6.7, showing the small ($< 10\%$) degradation in sensitivity when introducing the statistical uncertainties on background templates (preliminary results for a single mass point are provided). The same distributions and systematic uncertainties are considered as in the previous subsection. The ratio of the observed to the expected limit, along with median expected limit and the 1σ and 2σ fluctuations, for the m_H values considered are given in Table 6.6. A generally very good agreement is observed with the values from MCLIMIT reported in Table 6.5. The usage of the asymptotic approximation for the present analysis is expected to be well justified. However a direct comparison with the frequentistic prescription using ensembles of simulated measurements is presented in Table 6.8. As expected the differences in the two procedures are not significant. The 95% CL exclusion limits are presented in Figure 6.6 as a funcion of the Higgs mass hypothesis without considering template stat. uncertainties. Finally, Figure 6.7 represents the main result of this analysis and shows the obtained exclusion limits with the inclusion of all the systematic uncertainties.

m_H	Observed	Expected $\mu/\mu_{\rm SM}$				
(GeV)	$\mu/\mu_{\rm SM}$	-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
mH=120	63.0	35.0	46.9	65.1	95.6	141
mH=125	22.7	14.2	19.0	26.4	38.7	57.2
mH=130	11.8	6.1	8.2	11.4	16.6	24.5
mH=135	4.8	3.6	4.8	6.6	9.7	14.2
mH=140	3.7	3.1	4.1	5.7	8.3	12.1
mH=145	2.7	2.2	3.0	4.1	6.0	8.8
mH=150	3.6	2.3	3.1	4.3	6.2	9.0
mH=155	6.4	2.4	3.3	4.5	6.6	9.7
mH=160	7.6	4.8	6.4	8.9	12.9	18.9
mH=165	7.6	8.6	11.5	15.9	23.3	34.1
mH=170	8.8	9.3	12.5	17.3	25.2	36.9
mH=175	14.3	11.3	15.2	21.1	30.7	44.9
mH=180	16.9	11.8	15.8	21.9	31.9	46.6

Table 6.6 The observed 95% CL upper limits in the $H \to ZZ \to \ell\ell jj$ channel, as a multiple of the Standard Model rate, for an integrated luminosity of 4.71 fb^{-1}. The corresponding expected median upper limit, along with its $\pm 1\sigma$ and $+2\sigma$ values, is also shown.

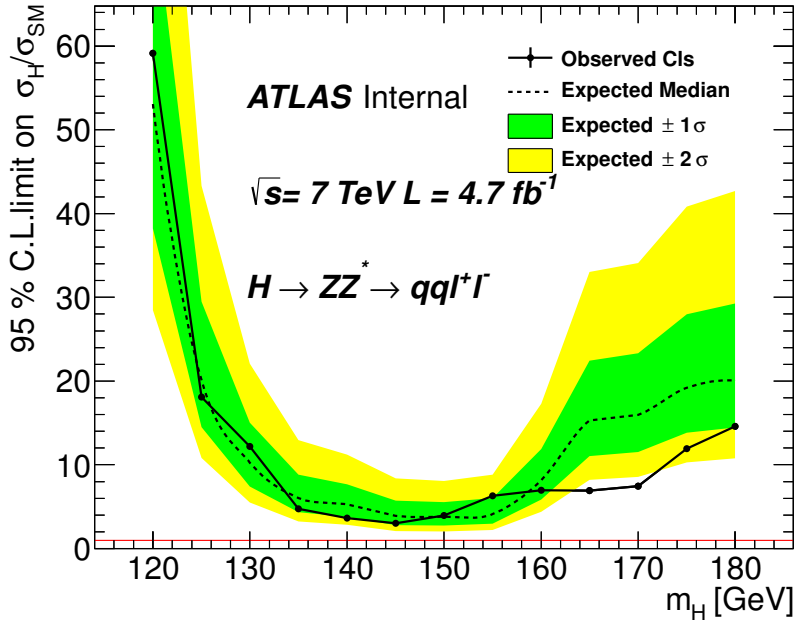
m_H (GeV)	Observed $\mu/\mu_{\rm SM}$	Expected $\mu/\mu_{\rm SM}$				
		-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
without stat. error	12.2	5.5	7.4	10.3	15.0	22.1
with stat. error	11.8	6.1	8.2	11.4	16.6	24.5

Table 6.7 The observed 95% CL upper limits in the $H \to ZZ \to \ell\ell jj$ channel, as a multiple of the Standard Model rate, for an integrated luminosity of 4.71 fb^{-1} for a 130 GeV Higgs mass hypothesis. Comparison between the case where bin-by-bin statistical uncertainties are considered or not considered in the limit setting procedure. The corresponding expected median upper limit, along with its $\pm1\sigma$ and $+2\sigma$ values, is also shown.

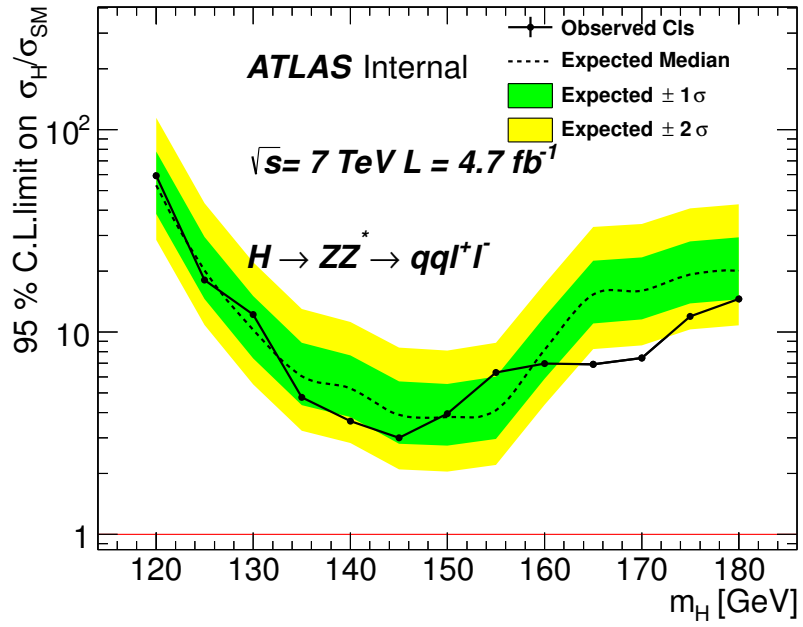
m_H (GeV)	Observed $\mu/\mu_{\rm SM}$	Expected $\mu/\mu_{\rm SM}$				
		-2σ	-1σ	Median	$+1\sigma$	$+2\sigma$
mH=140 asymptotic	3.6	2.8	3.8	5.3	7.7	11.2
mH=140 toys	–	–	–	–	–	–

Table 6.8 The observed 95% CL upper limits in the $H \to ZZ \to \ell\ell jj$ channel, as a multiple of the Standard Model rate, for an integrated luminosity of 4.71 fb^{-1} for a 140 GeV Higgs mass hypothesis. Comparison between the case where asymptotic approximation has been used or simulated measurements are thrown to derive the upper limits. The corresponding expected median upper limit, along with its $\pm1\sigma$ and $+2\sigma$ values, is also shown.

The expected and observed 95% CL upper limit on signal strength for various combinations of the individual channels are shown in Figure 6.8. In particular the combination of the untagged and tagged channel are shown for the electron, 6.8(a), and muon channels, 6.8(b), separately. The combination of electron and muon for the untagged, 6.8(c), and tagged,6.8(d), channels separately are also shown. In all cases a good agreement between the observed and expected limit is found with fluctuations below the 2 σ level. As can be seen electron, muon, tagged and untagged channels contribute with roughly equal sensitivity to the overall combination.



(a)



(b)

Figure 6.6 The expected (dashed line) and observed (solid line) upper limits on the total cross section divided by the expected Standard Model Higgs boson cross section, calculated using CL_s at 95%. Linear scale (a), log scale (b). The green and yellow bands, centred on the dotted line, indicate the range in which the limit is expected to lie in the absence of a signal. The red line shows the Standard Model value of 1. Results obtained without considering bin-by-bin statistical uncertainties on $m_{\ell\ell jj}$ templates.

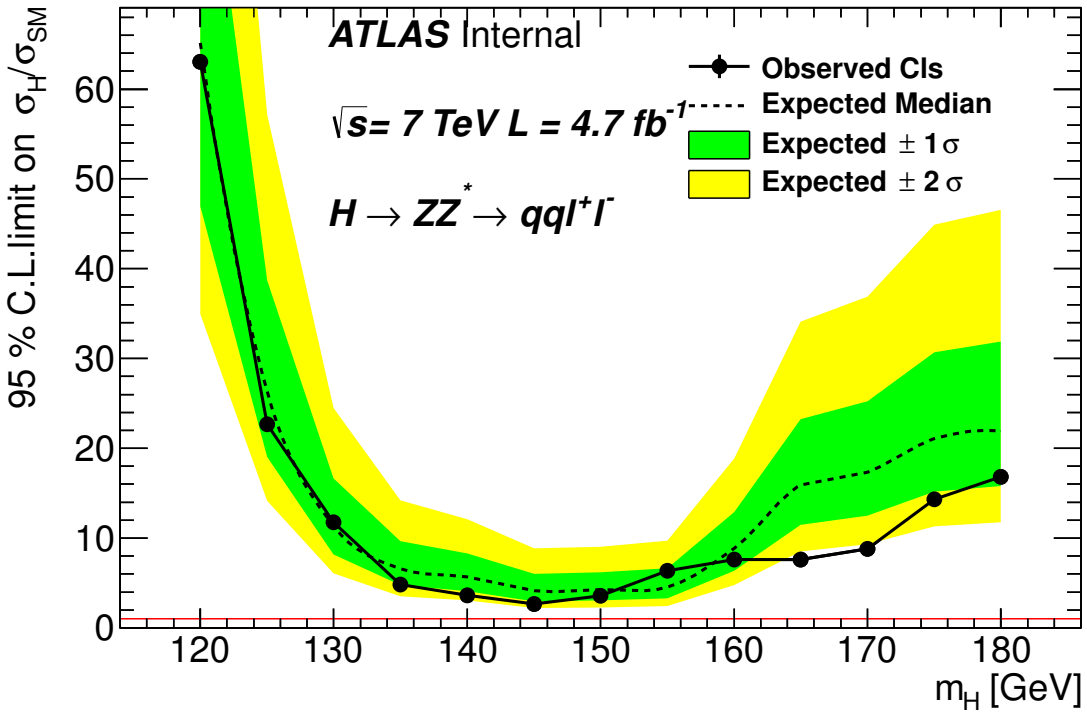
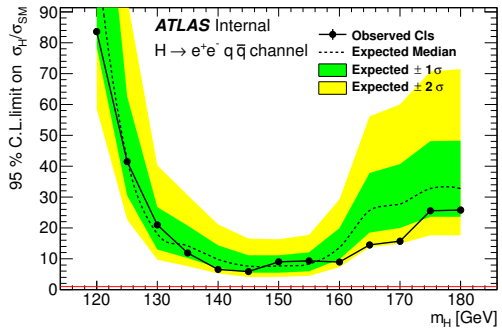
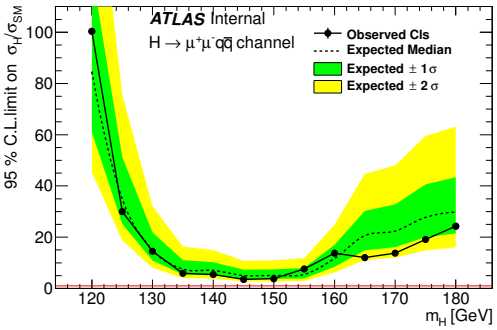


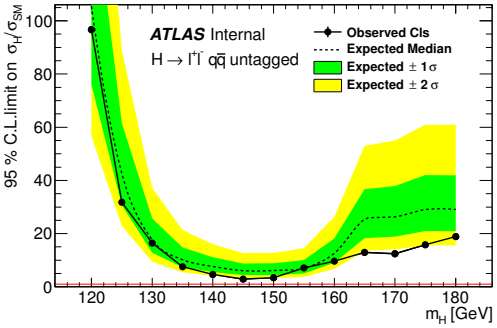
Figure 6.7 The expected (dashed line) and observed (solid line) upper limits on the total cross section divided by the expected Standard Model Higgs boson cross section, calculated using CL_s at 95%. The green and yellow bands, centred on the dotted line, indicate the range in which the limit is expected to lie in the absence of a signal. The red line shows the Standard Model value of 1.



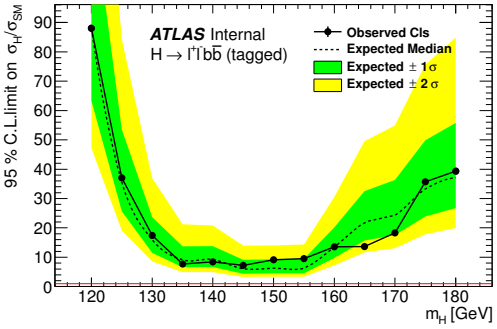
(a)



(b)

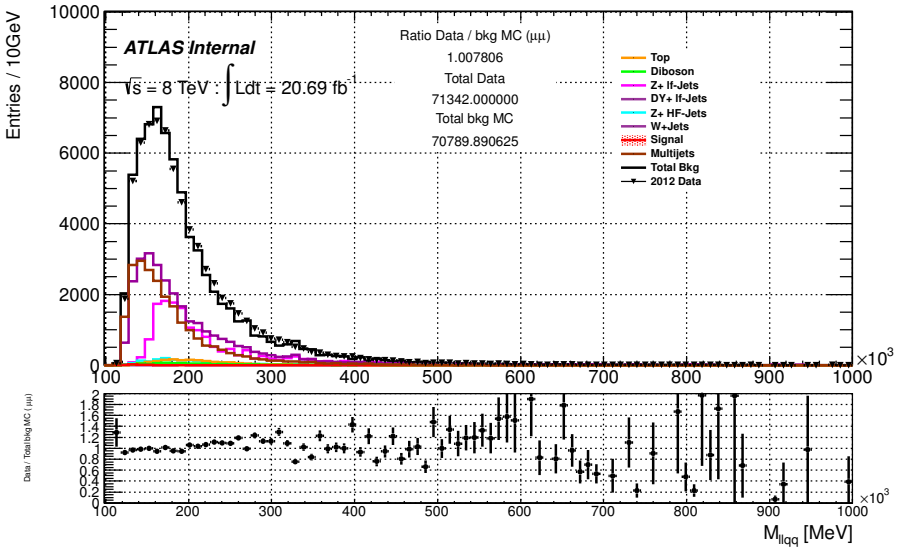


(c)

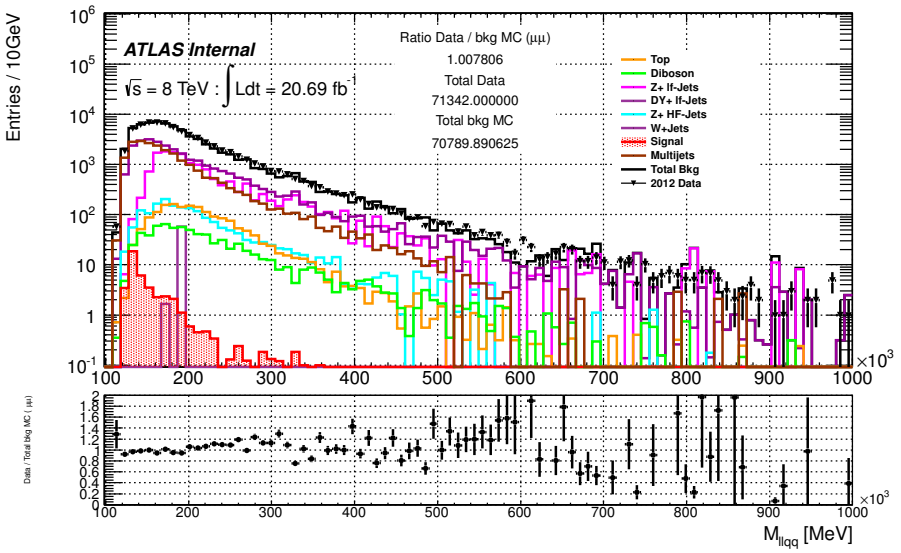


(d)

Figure 6.8 The expected (dashed line) and observed (solid line) upper limits on the total cross section divided by the expected Standard Model Higgs boson cross section, calculated using CL_s at 95%, for different combinations of input channels. The green and yellow bands, centred on the dotted line, indicate the range in which the limit is expected to lie in the absence of a signal. The red line shows the Standard Model value of 1. Electron channel only (a); muon channels only (b); untagged channels only(c); tagged channels only(d).

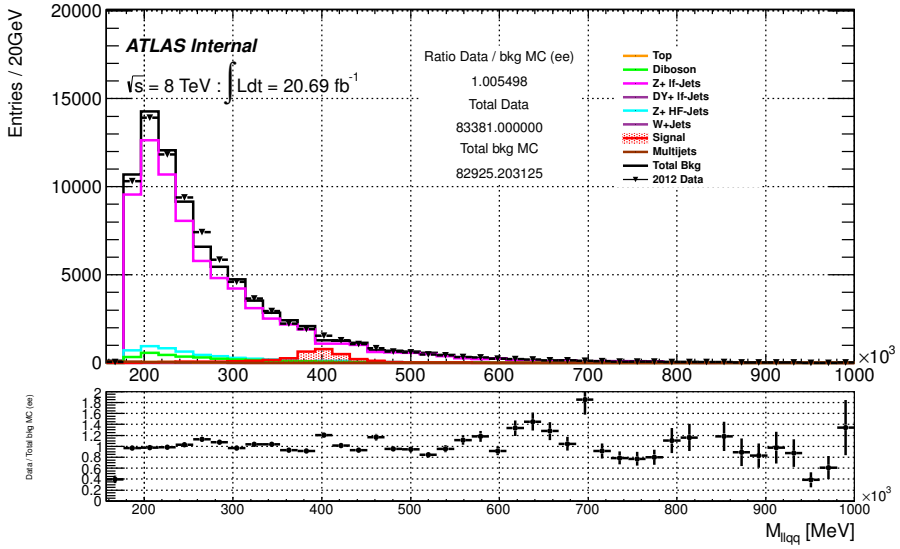


(a)

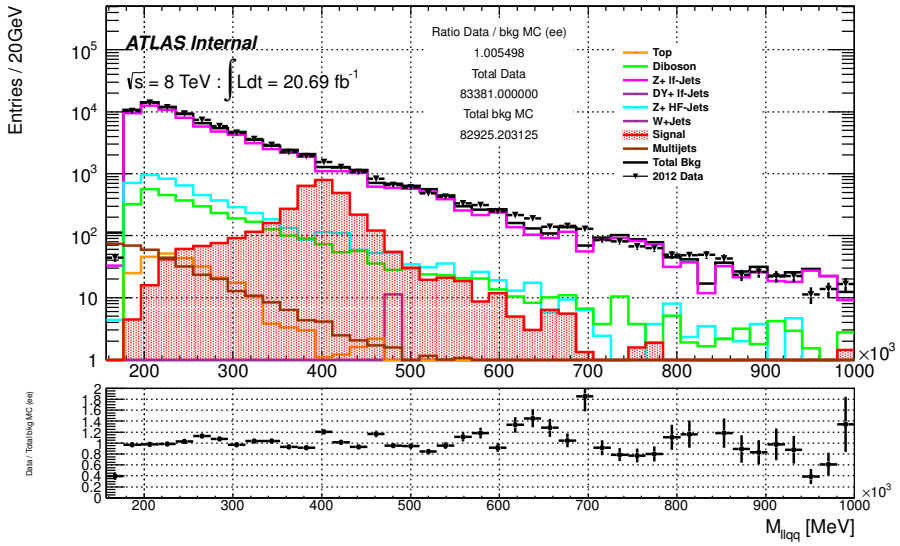


(b)

Figure 6.9 Preliminary M_{lljj} distribution for the muon channel in the case of the 2012 Low Mass (8 TeV) analysis with a Higgs mass hypothesis of 130 GeV *(under development)* in (a) linear and (b) logarithmic scale.



(a)



(b)

Figure 6.10 Preliminary M_{lljj} distribution for the electron channel in the case of the 2012 High Mass (8 TeV) analysis with a Higgs mass hypothesis of 400 GeV *(under development)* in (a) linear and (b) logarithmic scale.

Chapter 7

Optimization of Signal Significance: Quark-Gluon separation

The jets present into the signal process $H \to ZZ \to l^+l^-q\bar{q}$ always come from the hadronic decay into quarks of one of the Z boson. The main background process in such a analysis is the Z+jets, where the Z decay leptonically and is produced in association with jets. In this process the gluon radiation play an important role ($\approx 45\%$ of the jets in this background come from gluons).

In second place, the jet production is dominated from u or d quarks, those are inside the protons. The possibility to have −or build− discriminating variables relative to the flavour of the jet that can allows to distinguish the jet-origin can constitute an interesting tool to improve the experimental analysis.

The main difference comes from the fact the Higgs signal under study has not jets coming from gluons. Parallel to that it can be called the "standard" 2011 Low Mass analysis described into the previous chapter, was tested it under the untagged channel a series of Multivariable analysis tools in order to look for a method to discriminate the above mentioned jets using variables already identified by the ATLAS collaboration as possible discriminant due to their flavour-depending behaviour.

The possible improvement could come from the possibility to develop a tool be able to discriminate jets coming from light-quarks (u, d, s) of those coming from gluons. In ATLAS, using the jet reconstruction algorithm so-called TopoCluster (see Section 3.6), and thanks to the good calorimetric energy resolution is possible to measure the energy of the jet and its internal tracks distribution with a granularity that could allow to have the necessaries requests for the construction of a quark-gluon tagger.

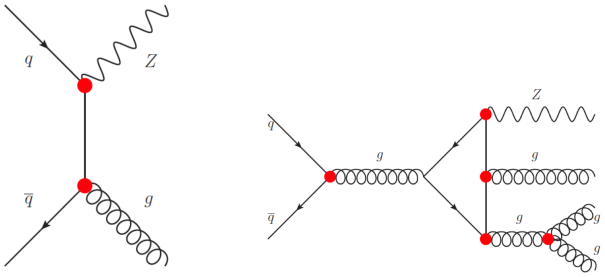


Figure 7.1 Feynman diagrams to exemplify the gluon production in Z+jets backgound.

7.1 Quark-gluon multivariate discriminant

The spin and the color-charge induce different experimental properties for quark and gluon jets. Leading order calculations foresee a higher track multiplicity for gluon-jets with respect to quark-jets[142] the expected ratio for the mean number of tracks is 2.25 (see Figure 7.2). Similar calculations have been applied also on angular variables and they show that gluon jets are wider than quark-jets (see Figure 7.3). Moreover LEP results [149] show that b-jets properties are closer to gluon-jets even if this similarity is expected to be less pronounced at high p_T.

Through an exhaustive search of existing and novel jet substructure observables it has been shown that a combination of two simple variables, the charge track multiplicity and the p_T weighted linear radial moment can filter out the greater part of the gluon jets while keeping more than half of the light-quark jets[151].

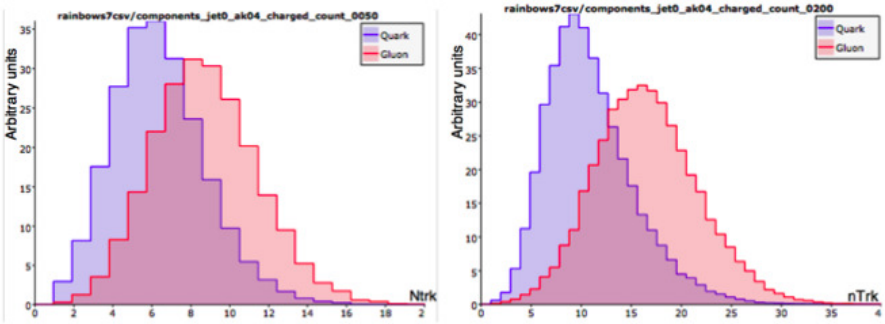


Figure 7.2 Number of tracks for gluon (red) and light quark (blue) jets for jet p_T of 50 GeV (left) and 200 GeV (right).

In ATLAS two similar quantities have been used: the number of charged tracks inside 0.4 cone around the jets (**nTrk**) and the jet width (**Width**) defined as:

$$Width = \frac{\sum_i p_{Ti} \Delta R_i}{\sum_i p_{Ti}} \tag{7.1}$$

where the sum is extended to all the tracks in 0.4 cone. The separation power of these two variables can be evaluated with MC dijets samples using standard likelihood ratio method using:

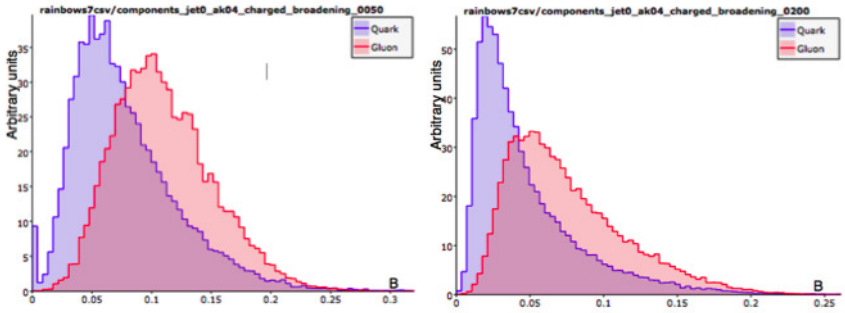


Figure 7.3 Broadening for gluon (red) and light quark (blue) jets for jet p_T of 50 GeV (left) and 200 GeV (right).

$$R = \frac{L_s(i)}{L_s(i) + L_b(i)} \tag{7.2}$$

where $L_s(i)$ and $L_b(i)$ represent the likelihood for the signal (light quark jets) and background (gluon jets) obtained using the pdf for the two variables (nTrk and Width). The flavor of the jets has been assigned considering the flavor of the most energetic parton inside the jet. As shown in Figure 7.4 the separation power between light quark and gluon jets depends mainly on the jet p_T suggesting good separation for $p_T \geq 50$ GeV. In the low m_H range, the p_T spectra of the jets coming from the hadronic decays of the Z has a large contribution in the low p_T regime so the quark-gluon tagging has been extended for jets up to 20 GeV.

A combined likelihood has been applied for both jets on all the events passing the standard $H \to ZZ \to l^+l^-q\bar{q}$ selection for the untagged channel. Results for the muon channel are shown in Figure 7.5.

Using the combined likelihood is possible to reduce the background contribution but the global significance is not improved.
In this analysis a large part of the events show a mixing between different jet flavours (1 quark + 1 gluon jet, 1 quark + 1b-quark jet, etc.. See Figure 7.6) and for this class of events the separation power of the standard likelihood method has been found to be very poor. For this reason an alternative approach based on a self-organizing map has been used[150].
A self-organizing map (SOM) or self-organizing feature map is a type of artificial neural network that is trained using unsupervised learning to produce a low-dimensional (typically two-dimensional), discretized representation of the input space of the training samples, called *map*. Self-organizing maps are different from other artificial neural networks in the sense that they use a neighborhood function to preserve the topological properties of the input space.

The vector input is represented by the nTrk and Width of the two jets for all the events passing the standard selection thus leading to 4-dimensional input space (Figure 7.7) and 4x4 rectangular SOM maps with gaussian neighborhood have been used. The maps has been splitted in 5 different p_T bins and trained with MC dijets sample similar to the one used for the likelihood method.

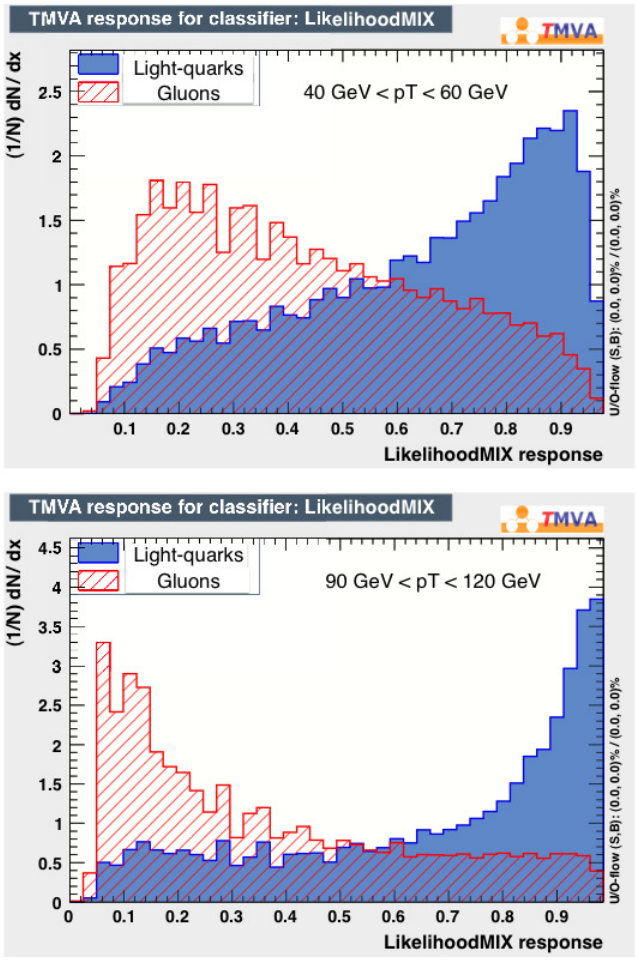


Figure 7.4 Likelihood ratio test results for $20 < p_T < 40$ GeV(top) and $90 < p_T < 120$ GeV(bottom) bins.

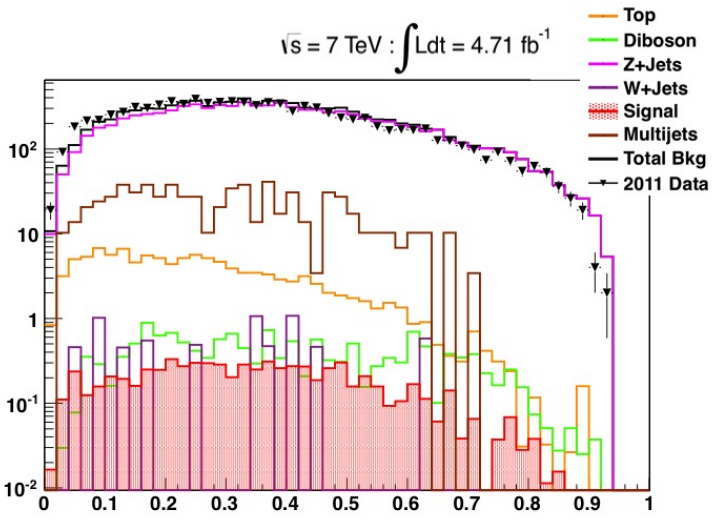


Figure 7.5 Combined likelihood for events passing the standard low mass selection (muons, untagged channel).

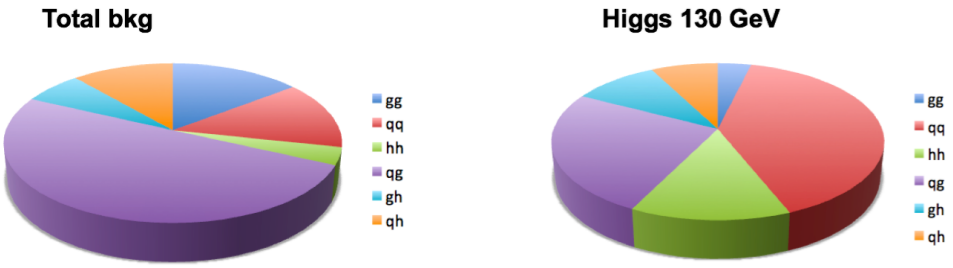
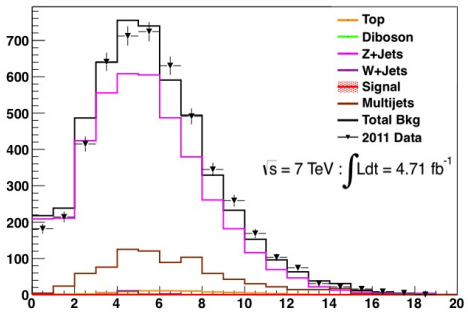
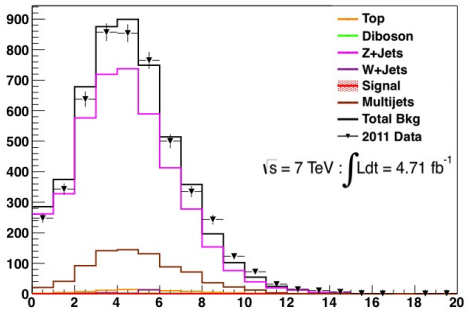


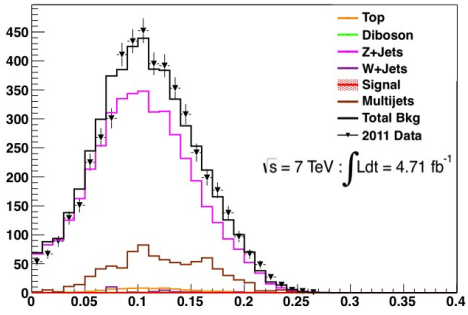
Figure 7.6 Different jet flavours composition for MC Higgs signal $m_H = 130$ GeV(left) and total background(right) after jet "truth" labelling.



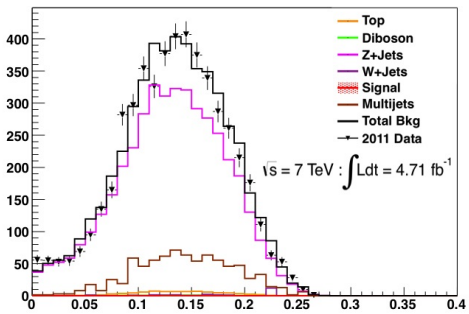
(a)



(b)



(c)



(d)

Figure 7.7 Ntrk (a) and (b), and Width (c) and (d) distributions for leading (left) and subleading jets (right) for all the events passing the standard low mass selection (electrons, untagged).

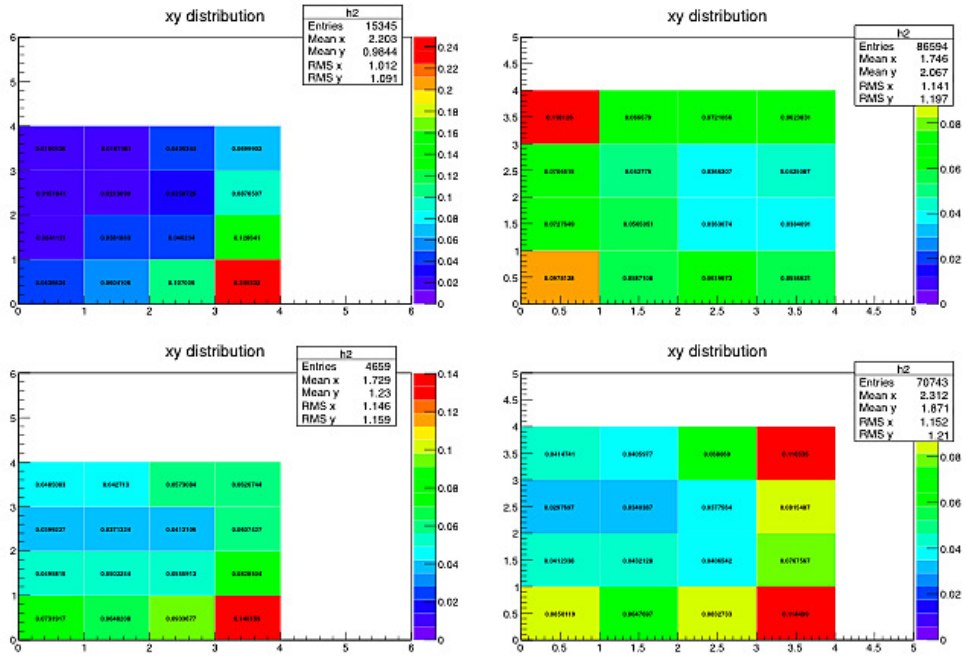


Figure 7.8 SOM maps response to MC dijets sample for the 60-90 GeV p_T bin: quark-quark jets (upper left), gluon-gluon jets (upper right), quark-gluon (lower left) and heavy flavor-heavy flavor jets (lower right).

Once the SOM maps have been defined the structure of the maps has been tested checking the response to well known input. As shown in Figure 7.8 the SOM neural network is able to group MC dijets events onto the 2D map on the basis of the dijet-flavor category.

7.2 Results in $H \to ZZ^* \to l^+l^-q\bar{q}$ Low Mass analysis 2011

Ones the possible "cuts" to be applied have been identified into the SOM output maps −always taking into account the reduction of background and the increase of the significance− the standard analysis (SA) is reproduced as describe the Chapter 5, adding such a SOM cuts (See Figure 7.9).

The distribution of the events passing the standard selection in the untagged channel using the SOM maps is shown in Figure 7.9. It is possible to recognize single neurons or cluster of contiguous neurons with large background contribution and poor or signal-free contribution. Removing these neurons from the SOM maps and hence from the selections contribute to the background reduction and enhance the $s/\sqrt{(b)}$ ratio.

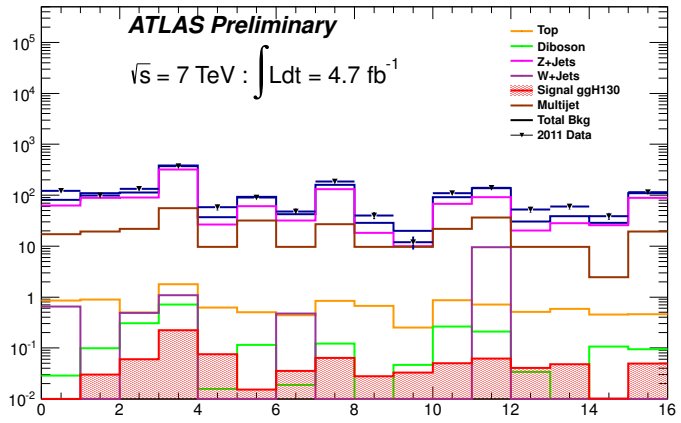
Significance Changes (2011 Low Mass analysis)		
Channel	PYTHIA Satandard Analisis (SA)	HERWIG++ SA + SOM
Electron gain	+5.7%	+6.5%
Muon gain	+0.1%	+0.7%

Table 7.1 Global significance changes for electron and muon Untagged channels. (ggH signal). The studies have been perform with two different MC training (PYTHIA and HERWIG++) samples in order to understand the possible systematic uncertainties.

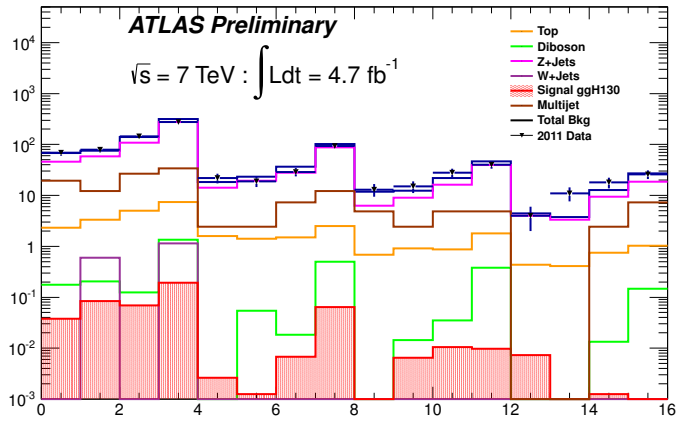
Total Background Reduction (2011 Low Mass analysis)	
Channel	after SOM
Electrons	20%
Muons	8%

Table 7.2 Total background reduction for electron and muon Untagged channels. Low Mass analysis 2011 dataset (ggH signal).

Bellow is showed some comparison plots: The m_{lljj} invariant mass before and after the application of SOM maps cuts is shown in Figure 7.10 for electron channel and in Figure 7.11 for muon channel. A conservative approach in SOM cuts definition leads to an overall background reduction of $\simeq 15\%$ (muons and electrons) and a global significance improvements of O(5% in the case of the electron channel).

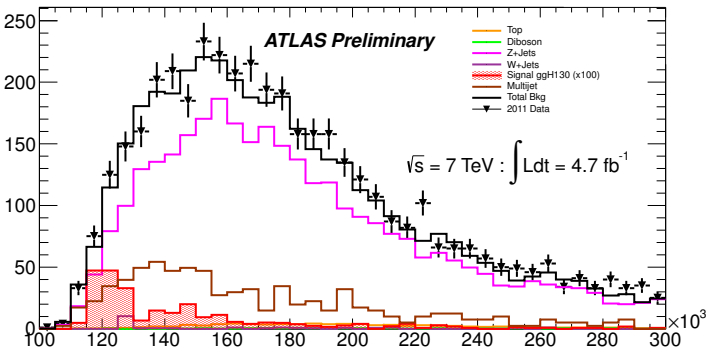


(a)

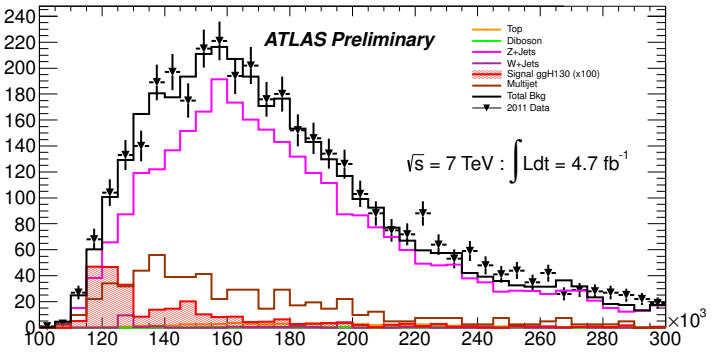


(b)

Figure 7.9 SOM maps for all the events passing the standard low mass selection (electrons, untagged) for the $20-40$ GeV (a) and $60-90$ GeV (b) p_T bin. 2D maps here has been shown as 1D segment with 16 (4x4) neurons.

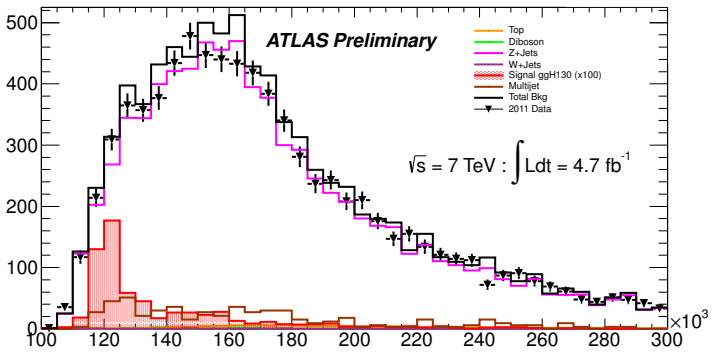


(a)

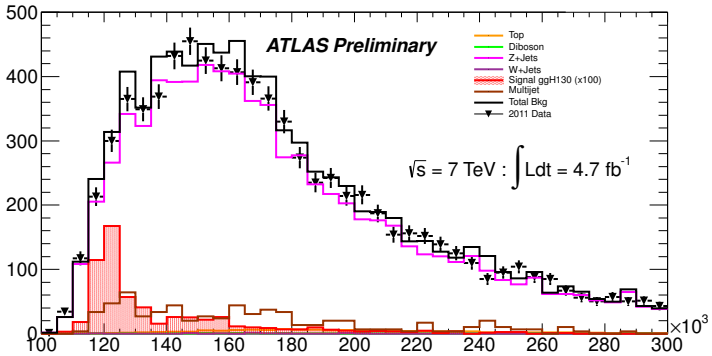


(b)

Figure 7.10 Comparison of data and MC for the M_{lljj} distribution in the untagged muon (a) and untagged electron channel (b) after the application of SOM cuts.



(a)



(b)

Figure 7.11 Comparison of data and MC for the M_{lljj} distribution in the untagged muon (a) and untagged electron channel (b) after the application of SOM cuts.

Chapter 8

Conclusions

Since the beginning, the search for the SM Higgs boson is one of the most crucial goals of the LHC scientific programme. Its high centre of mass energy enables not just to search for the SM Higgs boson at Low Mass regime i.e. 120 GeV $< m_H < 2 \times m_Z$, but to extend it to much larger masses, in the range of 200 GeV $< m_H <$ 1 TeV.

The discovery of this new particle has been made public at CERN on July 4th 2012, and it is based on the analysis and the combination of both the 2011 and part of the 2012 datasets, studied independently by the two collaborations (ATLAS and CMS). ATLAS reported a mass for this Higgs-like particle of $m_H \sim 125$ GeV. The mass, signal strength and couplings measurements have been updated using up to $4.8 fb^{-1}$ of pp collision data at $\sqrt{s} = 7$ TeV and about $21 fb^{-1}$ at $\sqrt{s} = 8$ TeV for the two channels $H \to \gamma\gamma$[26] $H \to ZZ^* \to 4l$[27].

Although a significant portion of this high mass range have been excluded at 95% CL, further possible extensions to the SM can conspire to allow a heavy Higgs boson to be compatible with existing measurements and resent Higgs boson-like particle discovery.

In this thesis the search for Higgs-like bosons in the $H \to ZZ^* \to \ell^+\ell^- q\bar{q}$ decay channel has been presented. The aim of this thesis was the description of all the Higgs-like search analyses performs by the ATLAS Collaboration using this final state.

The search for the Higgs boson has been perform in the Low Mass range using the $H \to ZZ^* \to \ell^+\ell^- q\bar{q}$ final state.

The sensitivity of this channel at Low Mass is not at the level of the $H \to ZZ^* \to 4l$ golden channel, however it has some specific properties that make this channel an excellent "test field" for new techniques and tools.

This analysis has been required an excellent knowledge and use of both leptonic (e and μ) and hadronic (jets and $b-$jets) objects. Moreover in the low mass regime the challenge has been increased due to the very low-p_T of the interesting objects. Leptons are well known objects in ATLAS. Here the presence of jets in the final state is one of the most difficult aspects of this analysis. The resolution of the reconstructed Higgs mass is worse and the requirement of at least two jets in the final states gives rise to high background contamination (see Figure 5.28) with $s/\sqrt{b} \sim 0.072$ in the m_{lljj} distribution for the muon channel and $s/\sqrt{b} \sim 0.045$ in the m_{lljj} distribution for the electron channel (see tables 6.1, 6.3).

This conditions have been taken as an opportunity to study some jet's properties about the possibility "separate" jets coming from the selected signal that those coming from background. Furthermore this is key point also for the High Mass analysis.

The search work started with the analysis using the data samples recorded by ATLAS in 2011 with a total luminosity of 4.7 fb^{-1}. The mentioned 2011 Low Mass analysis was concluded and approved by the ATLAS Collaboration[142]. This is a novel analysis not just for the mass range that was explored but because the new tools that have been developed in order to gain in significance and background rejection under the difficult environment of low p_T jets. Keeping in mind the fundamental idea of jet-understanding *or* jet-separation: one of the main improvements with respect to the previous analysis [113, 114] was the introduction of a criteria to select the jet-jet pair to build the hadronic Z boson candidate: the Kinematic Fitter (see Section 5.3.3).

Another relevant study is the inclusion of a tool that could allows to discriminate the flavour of the jet, improving significantly the rejection of the $Z/Drell-Yan+$jets background processes as the increase on the significance of the signal versus the main background: the Quark-Gluon tagging using SOM NN techniques (see Chapter 7).

Both tools: Kinematic Fitter and the Quark-Gluon tagging have been proven to help in the increasing of the significance of the signal over the background. Improvement have been obtained in the resolution of the reconstructed Higgs MC signal too.

This analysis has not the sensitivity to discover the 125 GeV Higgs-like particle into the 2011 total luminosity. After the full selection is applied to the 4.7 fb^{-1} of data recorded with the ATLAS detector in the 2011 LHC run, no evidence of the SM Higgs boson in this final state is found. For that reason, upper limits on its production cross section have been set: the best sensitivity is reached at $m_H = 145$ GeV where the expected limit on the SM Higgs production cross section is $4.1 \times \sigma_{\mathrm{SM}}$, and the observed limit is $3.4 \times \sigma_{\mathrm{SM}}$, while for $m_H \sim 125$ GeV, that is the mass of the discovered Higgs-like boson, the sensitivity of this analysis is $26.2 \times \sigma_{\mathrm{SM}}$.

At the same time, this thesis presented the actual status of the $H \to ZZ \to \ell^+\ell^- q\bar{q}$ analyses performed by ATLAS, where a leading role have been taken into the relevant collaboration physics group.

The aim is to perform a Higgs-like boson search in the *extended* mass range starting the 2012 Low Mass analysis −taking as baseline the already approved 2011 Low Mass analysis− and continue to the High Mass range (200 GeV $< m_H <$ 1 TeV).

To analyse the 2012 data sample many re-optimization of the analysis selections are needed, since it offers different challenges with respect to the 2011 analysis. This extension into the high masses regimes to search for new particles beyond the 600 GeV limit shows some theoretical problems concerning the interference of several processes giving similar final states that need to be taken into account. Many models beyond the Standard Model [29, 30] foresee new particles with very high masses which may decay to Z boson pairs. Therefore being able to efficiently select this final state is of crucial importance.

It has been proven that the χ^2-based jet selection criterion has a good performance in terms of significance and resolution on the $m_{\ell\ell jj}$ variable. Because of this, the physics group is actually studying the Kinematic Fitter and Quark-Gluon performance in the 2012 analysis of the full mass spectrum.

The preliminary results for the 2012 analyses (Low and High Mass) using the $20.7 fb^{-1}$ data collected have been presented and described in Chapter 5. Results show a reasonable agreement Data/MC, taking into account the preliminary stage of such studies and the still ongoing optimizations that have to be performed.

The final goal from the point of view of the data available (since at the beginning of the 2013 year, LHC starts a two years of technical stop) is to perform a 2011+2012 analysis in the all mass range (120 GeV − 1 TeV), looking not just for the best possible exclusion limits for the SM Higgs boson, but to search for possible new physics in this exciting $\sim TeV$-mass scale that LHC just starts to look over.

- "Mamita, estoy cansado!"
- "Ah!, dele, que ya falta menos que cuando empezó"
diálogo permanente.

Bibliography

[1] S. L. Glashow, *Partial-symmetry of weak interactions*, Nucl. Phys. **22** no. 4, (1961) 579.

[2] S. Weinberg, *A model of leptons*, Phys. Rev. Lett. **19**, (1967) 1264.

[3] A. Salam, *Weak and electromagnetic interactions*, in *Elementary particle theory: relaticistic groups and analyticity*, NN. Svartholm, ed., p.367, Almqvist & Wiksell, 1968, Proceeding of the eighth Nobel Symposium.

[4] G. Hooft and M. Veltman, *Regularization and Renormalization of Gauge fields*, Nucl. Phys. **B44**, (1972) 189.

[5] Particle Data Group Collaboration, K. Nakamura et al., Review of particle physics, J. Phys. G**37** (2010) 075021. 10, 54, 58, 101.

[6] F. Halzen and A. D. Martin, *Quarks and Leptons: An Introductory Course in Modern Particle Physics*. John Wiley and Sons, (1984) 11.

[7] P. Dirac, *The Quantum Theory of the Electron*, Proc. Royal. Soc. A**117** (1928) 610624

[8] C.N. Yang and R.L. Mills, *Conservation of isotopic spin and isotopic gauge invariance*, Phys. Rev. **96** (1954) 191-195

[9] N. Cabibbo, Unitary Symmetry and Leptonic Decays, Phys. Rev. Lett. **10** (1963) 531533.

[10] Peter W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*. Physical Review Letters **13** (1964).

[11] F. Englert, R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*. Physical Review Letters **13** (1964).

[12] C. R. Hagen G. S. Guralnik, T. W. B. Kibble, *Global Conservation Laws and Massless Particles*. Physical Review Letters **13** (1964).

[13] Gerald S. Guarlnik. *The History of the Guralnik, Hagen and Kibble development of the Theory of Spontaneous Symmetry Breaking and Gauge Particles*. International Journal of Modern Physics A**24** (2009).

[14] J. Goldstone, *Field theories with superconductor solutions. Il Nuovo Cimento* (1955-1965),**19** 154-164, 1961. 10.1007/BF02812722.

[15] Y. Nambu. *Quasi-particles and gauge invariance in the theory of superconductivity.* Phys. Rev., **117**. 648-663 (1960).

[16] J. Goldstone, A. Salam, and S. Weinberg. *Broken symmetries.* Phys. Rev., **127**. 965-970 (1962).

[17] T. Hambye, K. Riesselmann, *Matching conditions and Higgs boson mass upper bounds reexamined*, Phys. Rev. D**55** (Jun, 1997) 7255-7262.

[18] H. Georgi at al., Phys. Rev. Lett **40** (1978) 692; S.L. Glashow, D.V Nanopoulos, A. Yildiz, Phys. Rev. D**18** (1978) 196; K. Hikasa, Phys. Lett. B**164** (1985) 341.

[19] J.F. Gunion et al., Phys. Rev. D**34** (1986) 101; J. Gunion, G. Kane, J. Wudka, Nucl. Phys. B**299** (1988) 231; M. Dittmar, H. Dreiner, Phys. Rev. D**55** (1997) 167.

[20] A. Duperrin, arXiv:0805.3624 [hep-ph].

[21] V. Barger et all., Phys. Rev. D**44** (1991) 1426; D. Rainwater, D. Zeppenfeld JHEP 9712 (1997) 5.

[22] *CMS* Collaboration, Physics TDR, CERN/LHCC/2006-021 (2006).

[23] **ATLAS** Collaboration, Phys. Rev. B **716** (2012)**ATLAS**, *Collaboration, Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B**716** (2012) 1-29, arXiv:1207.7214 [hep-ex].

[24] **CMS** Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B**716** (2012) 30-61, arXiv:1207.7235 [hep-ex].

[25] **ATLAS** Collaboration, *Updated ATLAS results on the signal strength of the Higgs-like boson for decays into WW and heavy fermion final states*, ATLAS-CONF-2012-162 (2012).

[26] **ATLAS** Collaboration,*Higgs to Diphoton*, ATLAS-CONF-2013-012 (2013).

[27] **ATLAS** Collaboration, *SM H to $ZZ(*)$ to 4l*, ATLAS-CONF-2012-169 (2013).

[28] J. R. Ellis, G. L. Fogli, *New bounds on m(t) and first bounds on M(H) from precision electroweak data*, Phys. Lett. B**249** (1990) 543; J. R. Ellis, G. L. Fogli, E. Lisi, *Bounds on M(H) from electroweak radiative corrections*, Phys. Lett. B**274** (1992) 456 and *Indirect bounds on the Higgs boson mass from precision electroweak data*, Phys. Lett. B**318** (1993) 148.

[29] M. Kober, B. Koch and M. Bleicher, Phys. Rev. D**76** 125001 (2007).

[30] L. Randall and R. Sundrum, Phys. Rev. D**83** 4690 (1999).

[31] G.C. Branco, P.M. Ferreira, L. Lavoura, M.N. Rebelo, Marc Sher, Joao P. Silva, *Theory and phenomenology of two-Higgs-doublet models* Phys.Rept. **516** (2012) 1-102.

[32] B. Patt, F. Wilczek, *Higgs-field portal into hidden sectors*. May 2006. MIT-CTP-3745. e-Print: hep-ph/0605188.

[33] P. Ramond, Phys. Rev. D**3** (1971) 2415.

[34] A. Neveu, J.H. Scwarz, Nucl. Phys. B**31** (1971) 86; J.L. Gervais, B. Sakita, Nucl. Phys. B**34** (1971) 632.

[35] Torbjorn Sjostrand, Stephen Mrenna, Peter Z. Skands, *PYTHIA 6.4 Physics and Manual*. JHEP, 0605:026, (2006).

[36] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, et al. *HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes)*. JHEP, 0101:010, (2001).

[37] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, et al. *HERWIG 6.5 release note*. (2002).

[38] J.M. Butterworth, Jeffrey R. Forshaw, M.H. Seymour. *Multiparton interactions in photoproduction at HERA*. *Z.Phys.*, C**72**:637 646, (1996).

[39] P. Calafiura, W. Lavrijsen, C. Leggett, M. Marino, D. Quarrie. *The athena control framework in production, new developments and lessons learned*. Computing in High-Energy Physics (CHEP '04), (2005) 456-458.

[40] M. Bahr, S. Gieseke, M.A. Gigg, D. Grellscheid, K. Hamilton, et al. *Herwig++ Physics and Manual*. Eur.Phys.J., C**58** (2008) 639-707.

[41] M. Bahr, S. Gieseke, M. H. Seymour, *Simulation of multiple partonic interactions in Herwig++*. Journal of High Energy Physics, (2008) 7:76.

[42] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A.D. Polosa. *ALPGEN, a generator for hard multiparton processes in hadronic collisions*. JHEP, 0307:001 (2003).

[43] S. Frixione, B.R. Webber. *Matching NLO QCD computations and parton shower simulations*. JHEP, 0206:029 (2002).

[44] S. Frixione, P. Nason, B.R. Webber. *Matching NLO QCD and parton showers in heavy flavor production*. JHEP, 0308:007 (2003).

[45] S. Frixione, P. Nason, C. Oleari. *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*. JHEP, 0711:070 (2007).

[46] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig, E. Moyse, A. Salzburger. *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*. Technical Report ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002, CERN, Geneva, (2007).

[47] R. e. Kalman, *A new approach to linear filtering and prediction problems*. Journal of Basic Engineering 82 (1): 35-45. (1960). Retrieved 2008-05-03.

[48] **ATLAS** Collaboration. *Performance of the ATLAS Inner Detector Track and Vertex Reconstruction in the High Pile-Up LHC Environment.* Technical Report ATLAS-CONF-2012-042, CERN, Geneva, (2012).

[49] **ATLAS** Collaboration. *Performance of primary vertex reconstruction in proton-proton collisions at* $\sqrt{s} = 7$ TeV *in the ATLAS experiment.* Technical Report ATLAS-CONF-2010-069, CERN, Geneva, (2010).

[50] Georges Aad et al. *Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data.* Eur.Phys.J., C**72**:1909, (2012).

[51] **ATLAS** Collaboration, *Detector and Physics Technical Design Report*, Vol.1, CERN-LHCC-99-14 (1999).

[52] G. Aad et al. *Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data*, Eur.Phys.J., C**72**:1909, (2012).

[53] M. Aharrouche et al. *Measurement of the response of the ATLAS liquid argon barrel calorimeter to electrons at the 2004 combined test-beam. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 614(3):400-432, (2010).

[54] E. Abat, J. M. Abdallah, T. N. Addy, P. Adragna, M. Aharrouche, A. Ahmad, T. P. A. Akesson, M. Aleksa, C. Alexa, K. Anderson, et al. *Combined performance studies for electrons at the 2004 ATLAS combined test-beam.* Journal of Instrumentation, 5:11006, November (2010).

[55] **ATLAS** Collaboration, *Expected electron performance in the ATLAS experiment*, Tech. Rep. ATL-PHYS-PUB-2011-006, CERN, Geneva, (2011).

[56] B. Acharya et al. *Object selection and calibration, background estimations and MC samples for the Winter 2012 Top Quark analyses with 2011 data.* Technical Report ATL-COM-PHYS-2012-224, CERN, Geneva, (2012).

[57] **ATLAS** Collaboration, *Electron performance measurements with the ATLAS detector using the 2010 LHC proton-proton collision data.* arXiv:1110.3174 (2012, v2).

[58] **ATLAS** Collaboration, *Jet energy measurement with the ATLAS detector in proton-proton collisions at* $\sqrt{s} = 7$ TeV CERN-PH-EP-2011-191. Submitted to Eur. Phys. J. C.

[59] M. Cacciari, G.P. Salam, G. Soyez. *The Anti−k_t jet clustering algorithm.* JHEP, 0804:063 (2008).

[60] Stephen D. Ellis, Davison E. Soper. *Successive combination jet algorithm for hadron collisions.* Phys.Rev., D**48**:3160-3166 (1993).

[61] M. Wobisch, T. Wengler. *Hadronization corrections to jet cross-sections in deep inelastic scattering.* arXiv:hep-ph/9907280v1 (1998).

[62] **ATLAS** Collaboration. *Jet energy resolution and selection efficiency relative to track jets from in-situ techniques with the ATLAS Detector Using Proton-Proton Collisions at a Center of Mass Energy* $\sqrt{s} = 7$ TeV. Technical Report ATLAS-CONF-2010-054, CERN, Geneva, (2010).

[63] P. Adragna et al. *Testbeam studies of production modules of the ATLAS Tile Calorimeter*. Nucl. Instrum. Meth., A(606):362-394 (2009).

[64] LHC Higgs Cross Section Working Group Collaboration, S. Dittmaier et al., *Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables*, arXiv:1101.0593 [hep-ph].

[65] J. Pinfold et al. *Performance of the ATLAS liquid argon endcap calorimeter in the pseu- dorapidity region in beam tests*. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 593(3):324-342 (2008).

[66] **ATLAS** Collaboration. *Inputs to Jet Reconstruction and Calibration with the ATLAS De- tector Using Proton-Proton Collisions at* $\sqrt{s} = 900$ GeV. Technical Report ATLAS-CONF-2010-016, CERN, Geneva (2010).

[67] *ATLAS* Collaboration. *Commissioning of the ATLAS high-performance* $b-$*tagging algorithms in the* 7 TeV *collision data*. Technical Report ATLAS-CONF-2011-102, CERN, Geneva (2011).

[68] *ATLAS* Collaboration. *Measurement of the b-tag Efficiency in a Sample of Jets Containing Muons with* $5fb^{-1}$ *of Data from the ATLAS Detector*. Technical Report ATLAS-CONF-2012-043, CERN, Geneva (2012).

[69] *ATLAS* Collaboration. *Calibrating the* $b-$*Tag Efficiency and Mistag Rate in* $35pb^{-1}$ *of Data with the ATLAS Detector*. Technical Report ATLAS-CONF-2011-089, CERN, Geneva (2011).

[70] *ATLAS* Collaboration. *Measurement of the Mistag Rate with* $5fb^{-1}$ *of Data Collected by the ATLAS Detector*. Technical Report ATLAS-CONF-2012-040, CERN, Geneva (2012).

[71] *ATLAS* Collaboration. $b-$*jet tagging calibration on* $c-$*jets containing* D^{*+} *mesons*. Technical Report ATLAS-CONF-2012-039, CERN, Geneva (2012).

[72] Kenneth G. Wilson, *Confinement of Quarks*, Phys. Rev. D 10, 2445-2459 (1974).

[73] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. **13**, (1964) 321.

[74] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13**, (1964) 508.

[75] De Palma, M. et al., ALEPH: Techincal Report 1983 - CERN-LEPC-83-2.

[76] Bartl, W. et al., DELPHI: Techincal Proposal - DELPHI-83-66-1.

[77] von Dardel, G., Walenta, A. H., Lubelsmeyer, K., Deutschmann, M., Leiste, R., et al. L3 Technical Proposal. (1983). See the BOOKS subfile under the following call number: QCD197:E951.

[78] THE OPAL DETECTOR. TECHNICAL PROPOSAL - LEPC-83-4.

[79] Blair, R. et al. The CDF-II detector: Technical design report - FERMILAB-DESIGN-1996-01.

[80] Abazov, V. et al. D0 Run IIB upgrade technical design report (2002).

[81] The CDF and D0 collaborations, "Updated Combination of CDF and D0 Searches for Standard Model Higgs Boson Production with up to 10.0 fb-1 of Data", FERMILAB-CONF-12-318-E; CDF Note 10884; D0 Note 6348, arXiv:1207.0449.

[82] J. Beringer et al. (Particle Data Group), Phys. Rev. D*86*, 010001 (2012).

[83] **ATLAS** Collaboration, CERN/LHCC 97-21 (1997).

[84] **ATLAS** Collaboration, CERN/LHCC 97-19 (1997).

[85] **ATLAS** Collaboration, CERN/LHCC 97-20 (1997).

[86] **ATLAS** Collaboration, ATLAS Trigger Performances, ATL-COM-PHYS-2008-067, CERN (2008).

[87] S. Hassani, L. Chevalier, E. Lancon, J. F. Laporte, R. Nicolaidou and A. Ouraou, *A muon identification and combined reconstruction procedure for the ATLAS detector at the LHC using the (MUONBOY, STACO, MuTag) reconstruction packages*, Nucl. Instrum. Meth. A 572, 77 (2007).

[88] T. Lagouri et al., *A muon identification and combined reconstruction procedure for the ATLAS detector at the LHC at CERN*, IEEE Trans. Nucl. Sci. 51, 3030-3033 (2004).

[89] **ATLAS** Collaboration, *Selection of jets produced in proton-proton collisions with the ATLAS detector using 2011 data*, ATLAS-CONF-2012-020 (2012).

[90] **ATLAS** Collaboration, *Jet energy resolution and reconstruction efficiencies from in-situ techniques with the ATLAS Detector Using Proton-Proton Collisions at a Center of Mass Energy* $\sqrt{s} = 7$ TeV, ATLAS-CONF-2010-054 (2010).

[91] **ATLAS** Collaboration, *Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data*, ATLAS-CONF-2011-102 (2011).

[92] **ATLAS** Collaboration, *Measuring the b-tag efficiency in a top-pair sample with* 4.7 fb^{-1} *of data from the ATLAS detector*, ATLAS-CONF-2012-097 (2012).

[93] **ATLAS** Collaboration, *Performance of Missing Transverse Momentum Reconstruction in ATLAS with 2011 Proton-Proton Collisions at* $\sqrt{s} = 7$ TeV, ATLAS-CONF-2012-101 (2012).

[94] **ATLAS** Collaboration, *Jet energy resolution and reconstruction efficiencies from in-situ techniques with the ATLAS Detector Using Proton-Proton Collisions at a Center of Mass Energy* $\sqrt{s} = 7$ TeV, ATLAS-CONF-2010-054 (2010).

[95] J. Campbell, K. Ellis, C. Williams, *MCFM - Monte Carlo for FeMtobarn processes*, http://mcfm.fnal.gov/

[96] **ATLAS** Collaboration, *Measurement of the cross-section for b-jets produced in association with a Z boson at* $\sqrt{s} = 7$ TeV *with the ATLAS detector*, Phys.Lett.B 706 (2012) 295-313 (arXiv:1109.1403).

[97] GEANT4 Collaboration, S. Agostinelli et al., GEANT4: *A simulation toolkit*, Nucl. Instrum. Meth. **A 506** (2003) 250-303.

[98] **ATLAS** Collaboration, *The ATLAS simulation infrastructure*, Eur. Phys. J. **C 70** (2010) 823-874, (arXiv:1005.4568) [physics.ins-det].

[99] M. L. Mangano et al., *ALPGEN, a generator for hard multi-parton processes in hadronic collisions*, JHEP **07** (2003) 001.

[100] **ATLAS** Collaboration, *Measurement of the top quark pair production cross section with ATLAS in ppcollisions at* $\sqrt{s} = 7$ TeV *in dilepton final states*, ATLAS-CONF-2011-034 (2011)

[101] S. Frixione, B. R. Webber, *Matching NLO QCD Computations and Parton Shower Simulations*, JHEP **0206** (2002) 029 (arXiv:hep-ph/0204244).

[102] S. Frixione, P. Nason, C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070 (arXiv:0709.2092).

[103] S. Alioli, P. Nason, C. Oleari, and E. Re, *NLO Higgs boson production via gluon fusion matched with shower in POWHEG*, JHEP **04** (2009) 002 (arXiv:0812.0578).

[104] P. Nason and C. Oleari, *NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG*, JHEP **02** (2010) 037 (arXiv:0911.5299).

[105] E. Barberio and Z. Was, *PHOTOS: A Universal Monte Carlo for QED radiative corrections. Version 2.0*, Comp. Phys. Comm. **79** (1994) 291.

[106] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections*, arXiv:1101.0593 [hep-ph].

[107] J. M. Butterworth, J. R. Forshaw, and M. H. Seymour, *Multiparton interactions in photoproduction at HERA*, Z. Phys. **C72** (1996) 637-646 (arXiv:hep-ph/9601371).

[108] LHC Cross Section Working Group,
https : //twiki.cern.ch/twiki/bin/view/LHCPhysics/CrossSections
#CERN_Reports_Handbook_of_LHC_Hig

[109] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002 (2008).

[110] M. Cacciari, G. P. Salam, and G. Soyez, *Anti-KT jet clustering Algorithm*, JHEP **04** (2008) 063, (arXiv:0802.1189).

[111] G. Cowan, K. Cranmer, E. Gross, O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur.Phys.J.C71:1554,2011 (arXiv:1007.1727).

[112] L. Moneta, K. Belasco, K. Cranmer, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke, M. and Wolf, *The RooStats Project*, (2010) Available from: http://arxiv.org/abs/1009.1003, http://arxiv.org/abs/1009.1003 (arXiv:1009.1003).

[113] **ATLAS** Collaboration, *Search for a Standard Model Higgs boson in the mass range 200-600 GeV in the $H \to ZZ \to \ell^+\ell^- q\bar{q}$ decay channel with the ATLAS Detector*, ATLAS-CONF-2012-017 (2012).

[114] **ATLAS** Collaboration, *Search for a Standard Model Higgs boson in the mass range 200-600 GeV in the $H \to ZZ \to \ell^+\ell^- q\bar{q}$ decay channel*, CERN-PH-EP-2012-125 (arXiv:1206.2443).

[115] **ATLAS** Collaboration, Phys. Lett. B**710**, 49 (2012).

[116] **ATLAS** Collaboration, Phys. Rev. Lett. 108, 111803 (2012).

[117] **ATLAS** Collaboration, Phys. Lett. B**710**, 383 (2012).

[118] **ATLAS** Collaboration, Report No. ATLAS-CONF-2011-102, 2011, https://cdsweb.cern.ch/record/1369219/.

[119] **ATLAS** Collaboration, arXiv:1206.0756.

[120] **ATLAS** Collaboration, arXiv:1206.6074.

[121] **ATLAS** Collaboration, arXiv:1205.6744.

[122] A. L. Read, J. Phys. G 28, 2693 (2002).

[123] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, Eur. Phys. J. C**71**, 1554 (2011).

[124] R.D. Cousins, G.J. Feldman, Phys. Rev. D*57*, 3873 (1998)

[125] **CMS** Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B**716** (2012) 30âĂŞ61, arXiv:1207.7235 [hep-ex].

[126] **ATLAS** Collaboration, *Coupling properties of the new Higgs-like boson observed with the ATLAS detector at the LHC*, ATLAS-CONF-2012-127 (2012).

[127] **ATLAS** Collaboration, *Updated ATLAS results on the signal strength of the Higgs-like boson for decays into WW and heavy fermion final states*, ATLAS-CONF-2012-162 (2012).

[128] **ATLAS** Collaboration, *Combined coupling measurements of the Higgs-like boson with the ATLAS detector using up to* $25fb^{-1}$ *of proton-proton collision data* ATLAS-CONF-2013-034 (2013).

[129] https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TopMC10For2011Data.

[130] J. Butterworth, E. Dobson, U. Klein, B. Mellado Garcia, T. Nunnemann, J. Qian, D. Rebuzzi, R. Tanaka, *âĂćSingle Boson and Diboson Production Cross Sections in pp Collisions at* $\sqrt{s} = 7$ TeV, Tech. Rep. ATL-COM-PHYS-2010-695, CERN, Geneva (2010).

[131] https://twiki.cern.ch/twiki/bin/view/AtlasProtected/HforTool.

[132] B. Alvarez et al., *Measurement of Single Top-Quark Production in the Lepton+Jets Channel in pp Collisions at* $\sqrt{s} = 7$ TeV, Tech. Rep. ATL-COM-PHYS-2011-058, CERN, Geneva (2011).

[133] https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/HiggsZZllllSummer2012.

[134] **ATLAS** Collaboration, *Performance of Missing Transverse Momentum Reconstruction in ATLAS with 2011 Proton-Proton Collisions at* $\sqrt{s} = 7$ TeV. ATLAS-CONF-2012-101 (2012).

[135] https://twiki.cern.ch/twiki/bin/view/AtlasProtected/ExtendedPileupReweighting.

[136] N. Davidson, T. Przedzinski, Z. Was, *PHOTOS Interface in C++: Technical and Physics Documentation*, arXiv:1011.0937 [hep-ph].

[137] **ATLAS** Collaboration, *Muon reconstruction efficiency in reprocessed 2010 LHC proton- proton collision data recorded with the ATLAS detector*. Technical Report ATLAS-CONF-2011-063, CERN, Geneva (2011).

[138] **ATLAS** Collaboration, *Expected Performance of the ATLAS Experiment-Detector, Trigger and Physics*, arXiv:0901.0512 [hep-ex].

[139] **ATLAS** Collaboration, *Updated results and measurements of properties of the new Higgs-like particle in the four lepton decay channel with the ATLAS detector*. ATLAS-CONF-2012-169 (2012).

[140] **ATLAS** Collaboration, *Search for a heavy Standard Model Higgs boson in the mass range* $200 - 600$ GeV *in the channel* $H \to ZZ \to l^+l^-q\bar{q}\grave{I}\check{D}$ *with the ATLAS detector*, ATLAS-CONF-2012-017 (2012).

[141] **ATLAS** Collaboration, *Search for a heavy Standard Model Higgs boson in the mass range* $200 - 600$ GeV *in the channel* $H \to ZZ \to l^+l^-q\bar{q}$, ATLAS-COM-PHYS-2012-151 (2012).

[142] **ATLAS** Collaboration, *Study of the channel* $H \to Z^*Z \to \ell^+\ell^-q\bar{q}$ *in the mass range* $120 - 180$ GeV *with the ATLAS Detector at* $\sqrt{s} = 7$ *TeV* ATLAS-CONF-2012-163 (2012).

[143] **ATLAS** Collaboration, Jet/EtMiss Combined Performance Group, How to clean jets, https://twiki.cern.ch/twiki/bin/view/AtlasProtected/HowToCleanJets2011.

[144] **ATLAS** Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047 (2012).

[145] G. Cowan, K. Cranmer, E. Gross, O. Vitells, *Asymptotic formulae for likelihood-based tests of new physics*, Eur.Phys.J. C71 (2011) 1554, arXiv:1007.1727 [physics.data-an].

[146] A. L. Read, *Presentation of search results: The CL(s) technique*, J.Phys.G G28 (2002) 2693-2704.

[147] https://plone4.fnal.gov:4430/P0/phystat/packages/0711001; https://ppes8.physics.gla.ac.uk/twiki/bin/view/ATLAS/AtlasDataAnalysis

[148] **ROOT** Collaboration, *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*, CERN-OPEN-2012-016 (2012).

[149] A. et al Tech. Rep. CERN-PPE-95-126, CERN, Geneva (1995).

[150] T. Kohonen, *The self-organizing map*, Proceedings of the IEEE 78 (1990).

[151] J. Gallicchio, M. Schwartz, *Quark Gluon Tagging at LHC*, arXiv:1106.3076v2 [hep-ph].

[152] T. Lee, *A theory of spontaneous T violation*, Phys.Rev. D**8** (1973) 1226-1239.

[153] J. E. Kim, *Light pseudoscalars, particle physics and cosmology*, Phys. Rept. **150** (1987) 1-177.

[154] H. E. Haber, G. L. Kane, *The search for supersymmetry: probing physics beyond the Standard Model*, Phys. Rept. **117** (1985) 75-263.

[155] G. Branco, P. Ferreira, L. Lavoura, M. Rebelo, M. Sher, et al., *Theory and phenomenology of two-Higgs-doublet models*, Phys.Rept. **516** (2012) 1-102, arXiv:1106.0034 [hep-ph].

[156] A. Djouadi, *The Anatomy of electro-weak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model*, Phys.Rept. **459** (2008) 1-241, arXiv:hep-ph/0503173 [hep-ph].

[157] C.Y. Chen, S. Dawson, *Exploring two-Higgs-doublet models through Higgs production*, arXiv:1301.0309 [hep-ph].

[158] J. Chang, K. Cheung, P.Y. Tseng, T.C. Yuan, *Implications on the heavy CP-even Higgs boson from current Higgs data*, arXiv:1211.3849 [hep-ph].

[159] A. Drozd, B. Grzadkowski, J. F. Gunion, Y. Jiang, *Two-Higgs-doublet models and enhanced rates for a 125 GeV Higgs*, arXiv:1211.3580 [hep-ph].

[160] W. Altmannshofer, S. Gori, G. D. Kribs, *A minimal-flavor-violating 2HDM at the LHC*, Phys. Rev. D**86** (2012) 115009, arXiv:1210.2465 [hep-ph].

[161] D. S. Alves, P. J. Fox, N. J. Weiner, *Higgs signals in a type-I 2HDM or with a sister Higgs*, arXiv:1207.5499 [hep-ph].

[162] **CDF** Collaboration, T. Aaltonen et al., *Search for a two-Higgs-boson doublet using a simplified model in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV*, arXiv:1212.3837 [hep-ex].

[163] **CDF** Collaboration, T. Aaltonen et al., *Search for Higgs bosons predicted in two-Higgs-doublet models via decays to tau lepton pairs in* 1.96 TeV *p anti−p collisions*, Phys.Rev.Lett. **103** (2009) 201801, arXiv:0906.1014 [hep-ex].

[164] P. Ferreira, R. Santos, M. Sher, J. P. Silva, *Implications of the LHC two-photon signal for two-Higgs-doublet models*, Phys.Rev. D**85** (2012) 077703, arXiv:1112.3277 [hep-ph].